



## Durham E-Theses

---

### *Teachers' perceptions on the effectiveness of the Oxford Online Placement Test at King Abdulaziz University*

NASEM, HEBA, MOHAMMED, A

#### How to cite:

---

NASEM, HEBA, MOHAMMED, A (2019) *Teachers' perceptions on the effectiveness of the Oxford Online Placement Test at King Abdulaziz University*, Durham theses, Durham University. Available at Durham E-Theses Online: <http://etheses.dur.ac.uk/12989/>

#### Use policy

---

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in Durham E-Theses
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full Durham E-Theses policy](#) for further details.

**Teachers' perceptions on the effectiveness of the Oxford Online  
Placement Test at King Abdulaziz University**

**A thesis submitted to Durham University for the degree of Doctor of Philosophy**

**By**

**Heba Nasem**

**Durham University**

**School of Education**

**United Kingdom**

**2018**

## **Abstract**

English language placements tests are an essential component of preparatory year programmes (PYPs) as they serve to place students in an English course adequate to their level of proficiency. The aim of the study is to evaluate the perceived effectiveness of the Oxford Online Placement Test (OOPT), as seen by teachers in the English Language Institute (ELI) at King Abdulaziz University (KAU). The investigation explores teachers' views on the OOPT, on the ELI modules and on factors affecting students' test performance. It is framed by Messick's (1989) unified view of validity, which informed both the data analysis and the interpretation of the findings.

The study comprised three data collection stages: Stage 1: face-to face interviews, Stage 2: the questionnaire and Stage 3: telephone interviews. The majority of the questionnaire and interview respondents were unfamiliar with the OOPT, however, those who were familiar with the test, agreed that it was an effective tool and generally placed students in the correct level. However, some teachers felt that the content of the OOPT was not relevant to Saudi students.

Educational background, computer literacy and socio-economic status emerged as factors influencing achievement in the OOPT and ELI modules, according to teachers. Teachers also agreed that students specialising in the Sciences generally performed better compared to those specialising in the Arts.

With regards to the ELI modules, the majority of teachers felt that the courses were too short and that students who were placed in higher levels because of their OOPT scores were generally more proficient than those who had progressed through the ELI modules.

Implications for theory and practice are drawn from the findings as well as recommendations for the ELI and for future research.

## Table of content

<b>Abstract .....</b>	<b>ii</b>
<b>Table of content.....</b>	<b>iii</b>
<b>CHAPTER 1 Introduction.....</b>	<b>1</b>
1.1. Introduction .....	1
1.2. Context and rationale for the study.....	2
1.2.1. Personal narrative.....	3
1.3. Study objectives .....	5
1.4. Research questions.....	6
1.5. Significance of the study.....	6
1.6. Organisation of the thesis.....	7
<b>CHAPTER 2 CONTEXT OF THE STUDY .....</b>	<b>8</b>
2.1. Contextual background to the study .....	8
2.2. The Saudi educational system.....	9
2.3. The teaching of English in Saudi Arabia: An overview.....	10
2.3.1. Public and private schools .....	12
2.3.2. Overview of Higher Education in Saudi Arabia .....	14
2.4. King Abdulaziz University:.....	15
2.5. The preparatory year programme (PYP) at KAU.....	16
2.5.1. English courses within the PYP at KAU.....	17
2.6. Current admission and assessment procedures in the ELI .....	19
2.6.1. The OOPT at KAU .....	20
2.6.1.1. Content and structure of the OOPT .....	21
2.6.1.2. OOPT levels of proficiency vs. ELI levels.....	23
2.6.2. ELI achievement tests.....	26
2.7. The Cambridge Unlimited Special Edition Materials.....	27
2.8. Teachers' background.....	28
2.9. Students' background .....	29
2.10. Summary .....	29
<b>CHAPTER 3 LITERATURE REVIEW .....</b>	<b>30</b>
3.1. Introduction .....	30
3.2. The concept of validity over time.....	30
3.3. Messick's unified view of validity.....	33
3.4. Psychometric properties of tests .....	35
3.5. Washback: the effects of testing on teaching .....	35
3.6. Cultural appropriateness in testing and materials: .....	37
3.7. Assessment and testing.....	42
3.8. Types of tests and their purposes:.....	43
3.9. Placement or proficiency tests? .....	46
3.10. Placement tests.....	47



3.10.1. Purposes of placement tests in language programmes:.....	50
3.10.2. What do placement tests measure? .....	52
3.10.3. Challenges in placement testing:.....	52
3.10.4. The OOPT: .....	55
3.10.5. Relating test scores to the CEFR:.....	58
3.11. <i>Standardised tests and their relevance to the OOPT:</i> .....	61
3.12. <i>Computer Adaptive Tests (CATs):</i> .....	62
3.13. <i>TOEFL and IELTS tests:</i> .....	65
3.13.1. Institutional use of pre-arrival and placement tests for placement purposes:.....	66
3.14. <i>Teaching mixed-ability groups:</i> .....	68
3.15. <i>Reliability:</i> .....	69
3.16. <i>Test preparation:</i> .....	73
3.17. <i>Test familiarity:</i> .....	73
3.18. <i>Fairness and test bias:</i> .....	74
3.19. <i>Teachers' beliefs:</i> .....	76
3.19.1. What shapes teachers' beliefs? .....	77
3.20. <i>What affects students' test performance?</i> .....	78
3.20.1. Psychological factors .....	80
3.20.1.1. Motivation to learn English .....	81
3.20.1.2. Anxiety and stress .....	84
3.20.1.3. Experiential characteristics.....	85
3.20.2. Physical and physiological characteristics.....	86
3.21. <i>Summary</i> .....	86
<b>CHAPTER 4 RESEARCH METHODOLOGY</b> .....	88
4.1. <i>Introduction and research aims:</i> .....	88
4.2. <i>Research questions:</i> .....	89
4.3. <i>Mixed-methods research:</i> .....	90
4.4. <i>Methodological approach: case study:</i> .....	92
4.5. <i>Research design:</i> .....	93
4.6. <i>Data collection, research population and sampling:</i> .....	96
4.7. <i>Positionality and insider research:</i> .....	98
4.8. <i>Deciding on using synchronous or an asynchronous data collection:</i> .....	100
4.9. <i>Interviews: Stage 1 and Stage 3:</i> .....	101
4.9.1. Administering the interview:.....	104
4.10. <i>Questionnaire:</i> .....	106
4.10.1. Piloting the questionnaire: .....	108
4.10.2. Designing the questionnaire: .....	109
4.11. <i>Analysing the questionnaire and interview data</i> .....	110
4.12. <i>Reliability and validity</i> .....	113
4.12.1. Trustworthiness .....	114
4.12.2. Credibility:.....	115
4.13. <i>Ethical considerations:</i> .....	116
4.14. <i>Limitations of the chosen approach:</i> .....	118

4.15. Summary:.....	119
<b>CHAPTER 5 DATA ANALYSIS AND RESULTS .....</b>	<b>120</b>
5.1. Introduction:.....	120
5.2. Stage 1 – Analysis of face-to-face interviews:.....	123
5.2.1. Teacher’s views on the OOPT:.....	123
5.2.2. Students:.....	127
5.2.3. Modular system:.....	131
5.3. Stage 2 - The questionnaire: .....	133
5.3.1. Likert-type items: .....	135
5.3.2. Correlations: .....	142
5.3.3. Comparison between familiarity and unfamiliarity groups.....	145
5.3.4. Conclusions drawn from the quantitative analysis .....	147
5.4. Stage 3 – The telephone interviews .....	148
5.4.1. Teachers’ perceptions of the validity of the OOPT .....	148
5.4.1.1. Familiarity with the OOPT .....	149
5.4.1.2. Fairness.....	150
5.4.1.3. Effectiveness.....	155
5.4.1.4. The OOPT and the ELI curriculum .....	158
5.4.2. Perceived factors affecting students’ achievement.....	159
5.4.3. Drawbacks of the modular system.....	164
5.5. Summary.....	166
<b>CHAPTER 6 DISCUSSION .....</b>	<b>168</b>
6.1. Introduction .....	168
6.2. Validity of the OOPT for ELI students at KAU.....	171
6.2.1. Limited understanding of the OOPT .....	173
6.2.2. Cultural content of the OOPT .....	176
6.2.3. Skills tested in the OOPT .....	179
6.2.4. Computer literacy and administration of the test .....	181
6.3. Teachers’ beliefs and attitudes.....	181
6.4. Factors affecting students’ test performance.....	183
6.4.1. Previous learning experience .....	183
6.4.1.1. Equity between financial advantage and better opportunities.....	184
6.4.2. Extrinsic and intrinsic motivational factors.....	186
6.4.3. Student preparation .....	188
6.5. Teachers’ recommendations for practice.....	190
6.6. Areas for improvement to support students’ achievement.....	192
6.6.1. Teacher consistency across ELI modules.....	193
6.6.2. Addressing the strong reliance on textbooks in ELI modules.....	194
6.6.3. Employing an in-house placement test.....	195
6.7. Summary.....	196
<b>CHAPTER 7 CONCLUSION AND RECOMMENDATIONS .....</b>	<b>198</b>
7.1. Introduction .....	198
7.2. Revisiting the research questions .....	198

7.2.1. Research question 1: How do teachers view the OOPT? .....	199
7.2.2. Research question 2: How does the OOPT impact on placement and teaching, in the opinions of teachers? .....	200
7.2.3. Research question 3: What are the factors that might affect students' achievement in both the OOPT and during the English module, in the opinion of teachers? .....	201
7.2.4. Research question 4: How can the system be improved to support students' achievement? .....	202
7.3. <i>Other emerging findings</i> .....	202
7.4. <i>Contribution to knowledge</i> .....	203
7.5. <i>Contribution to theory</i> .....	204
7.6. <i>Recommendations and implications</i> .....	205
7.6.1. Recommendations drawn from findings related to the research questions .....	205
7.6.2. Recommendations not drawn from the research questions .....	207
7.6.3. Wider implications for teaching and learning .....	209
7.7. <i>Suggestions for future research</i> .....	211
7.8. <i>Limitations</i> .....	213
7.9. <i>Conclusion of the Thesis</i> .....	214
<b>Bibliography</b> .....	215
<b>Appendices</b> .....	253

**Acronyms and Abbreviations:**

AERA: American education research association

CAA: Computer-assisted assessment

CAT: Computer Adaptive test

CBT: Computer-based tests

CEA: Commission of English

CEFR: Common European Framework Reference for Languages

CEPT: Cambridge English placement test

CSC: Computer Science Course

EGP: English for General Purposes

ELI: English Language Institute

ESP: English for Specific Purposes

ETS: Educational Testing Service

GE: General Education

HE: Higher Education

IELTS: International English language testing system

KAU: King Abdulaziz University

KAUST: King Abdullah University of Science and Technology

KFPMU: King Fahad Petroleum and Mineral University

LPD: Language Policy Division

NCME: National council on measurement in education

OOPT: Oxford Online placement test

OUP: Oxford University Press

PYP: Preparatory year programme

SA: Study abroad

SES: Socio-economic status

SLA: Second language acquisition

SLO: Student learning outcomes

TL: Target lagunage

TC: Target culture

TOEFL: Test of English as a foreign language

UCQ: University of Calgary, in Qatar

## **Declaration**

This thesis and the work presented in it are the result of my own original research. No part of this work has been submitted for any other degree at this or any other university. Where other sources of information have been used, these have been acknowledged.

### **Statement of Copyright**

The copyright of this thesis rests with the author. No part of this thesis should be published, reproduced or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the author's prior consent and information derived from it should be acknowledged.

## **Acknowledgements**

I would like to express my gratitude to a number of key individuals who have provided their support and assistance during the course of my doctoral study. First, and foremost, I would like to express my sincere gratitude to my supervisor Prof. Peter Tymms, for his immense support and guidance from the early stages of my journey and till the end. I am also profoundly grateful to Dr. Nicola Reimann, for her continuous advice and encouragement during the later stages of my PhD. I have learnt a great deal from her constructive comments. My research ability has evolved as a result of their supervisions. I had the honour to work with two known academics.

Beside my supervisors, I would like to thank my family: my parents (Mohammad Nasem and Hind Halwani) my brothers (Loai, Nedal, Eyad, Hani, Mahmoud), and my sister (Fatma) who supported and motivated me throughout my PhD as well as my mother-in law (Mariam Al-Hammami) for her invaluable help and prayers.

Heartfelt thanks go to my dearest husband (Majed Badghaish), who supported me for the duration of my doctoral journey and created an educational environment at home. I am grateful to my two boys (Mohammad and Yousof) who have inspired me with their love and care and helped look after their younger siblings (Jana and Al-Badr) who also enlightened me in difficult moments.

I would like to extend my thanks to all my friends (Hadeel, Noha, Mona, Sara, Khadija, Enass and Ender) who contributed in to the successful completion of this study, without whom, the PhD journey would have been challenging.

My sincere thanks to Dr. Badia Hakim, the ELI dean of the women's campus, who facilitated my data collection. I am also thankful to the government of Saudi Arabia and the Ministry of Education for sponsoring and giving me the opportunity to complete my doctoral study here at Durham University.

## CHAPTER 1 Introduction

### 1.1. Introduction

Research into teachers' beliefs about assessment is a growing field globally (Barnes et. al., 2015, Fulmer et al., 2015), however little is known about how teachers conceptualise assessment within the context of Saudi Arabia. In particular, the use of placement tests in PYPs in Saudi Arabia remains under-researched and, to my knowledge, no previous study has explored teachers' beliefs on the effectiveness of the Oxford online placement test (OOPT) within this context. It has been previously acknowledged that 'the study of teachers' conceptions about assessment is a critical issue in the field of assessment research' Opre, 2015, p.229). However, while this field of research has implications for both policy and practice, little is known about teachers' conceptions regarding assessment (ibid.).

Furthermore, while studies which explored teaching in Higher Education (HE) have exposed the importance of academics' conception for their practices (e.g., Trigwell and Prosser, 1996), 'there is relatively little corresponding empirical research into how (HE) teachers understand assessment and their associated assessment practices' (Sadler and Reimann, 2018, p.131).

Understanding teachers' beliefs is central to gaining a more complete picture of the extensive number of factors, which can influence test performance. A number of studies have demonstrated a link between conceptions of teaching held by teachers and approaches adopted to teaching (e.g. Trigwell et al., 1994). Some studies also found a relationship between teachers' approaches to teaching and their students' approaches to learning (Trigwell et al., 1999). Teachers can also provide a personal perspective on factors, which, in their view, can affect test performance. Because teachers spend a significant amount of time with their students, they are able to observe students' learning over time and identify factors that appear to be associated with underperformance. Brown and Harris (2016, p.3) argue that 'test responses are a function not only of the items, tasks, or stimulus conditions, but of the *persons* responding and the *context* of measurement'. Aside from the structure and content of the test itself, equally important are 'factors in the environmental background' and the 'assessment setting' (Messick, 1989, p. 14). It has been previously suggested that 'student characteristics' as well as 'social and educational experiences' can influence performance (Messick, 1989, p. 215). This is relevant to the current study as the cultural



setting plays an important role in the investigation. Holliday (2016) advocates research that draws from opportunities arising in cultural and social settings. This kind of research can provide new perspectives on culturally-related factors that may influence test performance. It has been previously emphasised that culture and social context play a vital role in understanding educational systems, practice, forms of pedagogy and teachers' beliefs (Alexander, 2000; Andrews, 2007; Osborne, 2003 and Pepin, 2005).

While the study primarily aims to gain a better understanding of teachers' views on the OOPT, it is argued that the results can also provide data about the test's validity. According to Messick (1989), perceptions, uses and consequences of a test can also provide information about a test's validity (see chapter 3).

The investigation explores teachers' perceptions of the OOPT (e.g. whether or not they view it as an effective placement test), the complexity behind employing the OOPT within the cultural context (i.e. female students in a PYP at KAU) and the perceived consequences of administering this type of test (e.g. misplacement and implications of misplacement for teaching and learning). The context of the study is particularly relevant to the investigation as a number of concerns voiced in the teachers' interviews were closely linked to students' educational background and the socio-cultural context of the study.

## **1.2. Context and rationale for the study**

The government of Saudi Arabia has invested a significant amount of money over the years in order to improve the education system of the country (Ministry of Education, 2004). This is evident, for instance, from the Ten-Year Plan 2004-2014, released by the Saudi Ministry of Education (MOE), which included the development of infrastructure so that technology could be more easily employed in education (Ministry of Education, 2004). More recently, the teaching of English both in schools and in HE has gained greater importance and formed a central part of the curriculum. The importance of English is directly related to the country's interest in the global market and the fact that English has become the language of science, technology, business and commerce (urRahman and Alhaisoni, 2013). This has resulted in an increased focus on the teaching of English in HE, for example through mandatory PYPs.

The role of PYPs in ensuring graduates across all disciplines achieve a satisfactory competence in English is crucial to universities in Saudi Arabia and therefore research, which explores current practice within these programmes can make a significant contribution to knowledge.

### **1.2.1. Personal narrative**

Many institutions adopt placement tests to determine the appropriate English language level for students entering the PYP. The OOPT is one of placements tests employed in some of the universities in Saudi Arabia, including KAU. As I myself am a member of staff of the ELI, my own experience also informed the rationale for this study. In the four years prior to undertaking my PhD, I taught a range of courses in the ELI and thus personally experienced some of the issues voiced by my colleagues and students. It occurred to me that particularly in the lower levels there was a wide range of ability within one group and many students were very weak. In some of the informal discussions I had with colleagues in the ELI, concerns were raised regarding the large number of students being placed in the lowest ability levels. Some attributed some of these issues to the OOPT and assumed that the test was not placing students correctly. In particular they complained about the wide range of ability within a given level. Some colleagues commented about the weak computer-literacy skills of some students as well as unfamiliarity with online placement tests and felt that these factors disadvantaged students, thus resulting in lower marks. Colleagues often associated the weak computer-literacy skills of some students with their family background. They pointed out that some students came from rural areas and did not have a computer or internet access at home.

Some teachers also believed that in some cases students were placed in a level above their ability as a result of having correctly guessed a number of items on the test and having more secure test-taking skills compared to other students. Similar complaints were also voiced by students. Some argued that the level in which they were placed was either too high or too low for their ability. The cultural context of the ELI programme at KAU thus appeared to be linked to the way the OOPT was perceived. Some teachers were critical of the test because they felt it disadvantaged students with weak computer literacy, while others criticised it for its topics, which, in their view, were not culturally relevant for Saudi students.

These informal discussions with colleagues prompted me to explore the issue further and provided a starting point for my doctoral research. It appeared that the OOPT, while compulsory for all students, was not linked to the ELI curriculum, nor had anyone explored issues of misplacement, which could have resulted from the test or interpretation of test results. I felt that the OOPT was perceived as external to the ELI's PYP and thus issues relating to its effectiveness may have been overlooked. My own experiences, as well as some of my colleagues' comments, made me reflect upon the appropriateness of the OOPT for this specific context. Most students sitting the placement test would have had little or no exposure to authentic language use because of the approaches adopted for the teaching of English in Saudi public schools. Because of students' weak language skills upon entry I felt that students would not be able to understand idiomatic language use or more informal uses of English in conversations. I could relate to many of the concerns voiced by my colleagues as I also experienced similar issues in my own teaching. It was evident that the majority of students taking the OOPT were being placed in the two lowest levels, which made me question whether students' ability upon entry was indeed that weak or whether they were being misplaced as a result of low OOPT scores. The questions that arose through my teaching practice and conversations with colleagues who were raising similar issues prompted me to investigate this issue further. I was keen to explore whether students were indeed being misplaced as a result of the OOPT, whether there were other factors involved, and if so, what could be done to address this issue and improve the system both for teachers and students. It should also be pointed out that my role as a practitioner within the institution meant that I could undertake this research from an insider's perspective (see chapter 4).

The rationale for researching my own field of practice was not only informed by questions which arose from my own teaching, as well as similar concerns voiced by colleagues, but also from a wish to explore ways to improve practice within the ELI. Unfortunately, while students in my own classes expressed frustration with their OOPT results and some asked if it was possible to re-sit the exam as they felt that they had not been placed in the correct course, the head of placement tests and language coordinators spoke in defence of the OOPT and argued that it was a much better placement test compared to other tests available on the market. This brought me to reflect upon the importance of teachers' beliefs and perceptions

about testing, which is central to the investigation. I felt that teachers' views on the placement test and students' frustration with their OOPT scores did not appear to influence decision-making, hence one of the aims of the thesis is also to expose the importance of teachers' voices and involving teachers in discussions concerning assessment. It is argued that teachers' views should inform decision-making about testing as it directly affects their practice.

### **1.3. Study objectives**

Research into teachers' beliefs provides a means for understanding how beliefs may be related to student outcomes as well as providing insight into classroom practices and pedagogy (Kagan, 1992; Muis and Foy, 2010). Prior research has explored how teachers' background may influence practice, either directly or indirectly through the connection with knowledge, values and conceptions (Fulmer et al., 2015). While some studies did not find any significant differences according to these variables (e.g. Eren, 2010), several other studies have identified important relationships with background variables, such as subject matter or teaching experience (Fulmer et al., 2015, p. 483).

The study aims to evaluate teachers' beliefs of the OOPT as an effective tool and explore in greater depth their views on factors affecting students' performance both in the OOPT and in the ELI modules. Borg (2003) argues that teachers make decisions based on their practice, context, personal knowledge, thoughts, and beliefs. Understanding teachers' beliefs can therefore also help us understand the rationale behind the decisions they make, for instance with regards to teaching methodology and assessment. Jamalzadeh and Shahsavari (2015) argue that, in order to understand teachers' beliefs and practices, we need to have insight about the teaching context, teachers' needs and their methodologies.

While the term 'beliefs' is used here, the thesis also makes reference to other terms such as teachers' perceptions, views, conceptions and attitudes and these are used interchangeably.

As the study provides a teachers' perspective on the effectiveness of the OOPT as a placement test and their views on the ELI modules, understanding the importance of teachers' beliefs is central to the investigation.

The thesis is organised around the following five broad research objectives:

1. Identification of teachers' beliefs on the use of the OOPT within the ELI at KAU
2. Understanding what factors might affect students' performance, according to teachers.
3. Understanding teachers' views on the ELI modules so that the current system can be improved.
4. Understanding ways in which the OOPT may impact on teaching within the ELI
5. Determining ways in which the current curriculum can be improved in order to support students' learning

The study employs a mixed-methods research design consisting of three data collection stages: Stage 1: Face-to-face interviews, Stage 2: Questionnaire, Stage 3: Telephone interviews. The initial interview (Stage 1) served to inform the questionnaire design (Stage 2). The second interviews (Stage 3) served to follow up on the questionnaire and findings from the first interview.

#### **1.4. Research questions**

The study aims to investigate teachers' perceptions on the effectiveness of the OOPT in the PYP at KAU and it is guided by the following research questions:

1. How do teachers view the OOPT?
2. How does the OOPT impact on placement and teaching, in the opinions of teachers?
3. What are the factors that might affect students' achievement in both the OOPT and during the English module, in the opinion of teachers?
4. How can the system be improved to support students' achievement?

#### **1.5. Significance of the study**

The study provides a significant contribution to knowledge as little research has explored the use of English language placement tests in Saudi Arabia. Furthermore, the study explores teachers' perspective on a specific placement test and provides a perspective from the female section of the ELI, which makes the study unique among existing research on placement tests and PYPs in the country.

One of the challenges that both teachers and students face is the mismatch between the English instructions students received in secondary schools and the expectations at university level. Because schools focus primarily on teaching grammar and assess students accordingly, the majority of students enter university with very weak communicative skills in English. The OOPT, however, is aimed to test primarily students' understanding of authentic interaction in a communicative setting. A previous study on the use of an English language placement test in the ELI at KAU found that the majority of test takers are placed in the lowest ELI level (level 1) (Mahmoud, 2014), which could be linked to the type of skills tested in the placement test.

Understanding teachers' views on the factors which may influence students' performance in the OOPT as well as in the ELI modules is paramount to informing current practice and ensuring the ELI modules can effectively support students' learning throughout the PYP.

## **1.6. Organisation of the thesis**

The thesis opens with the introduction followed by Chapter 2, which contextualises the study within its socio-cultural and geographical context. This chapter describes HE in Saudi Arabia as well as the curriculum adopted in the ELI at KAU. The chapter outlines how the OOPT is employed to place students in the four English language modules of the PYP. This is followed by Chapter 3, which begins by laying out a critical review of the literature on testing (placing particular emphasis on placement tests), teachers' beliefs and factors affecting students' test performance. Chapter 3 further illustrates how the literature has informed the present study and refined the research aims guiding the investigation. Chapter 4, the methodology, describes the nature of the research as well as the rationale for adopting a mixed-methods design. This chapter also describes the sample and research instruments used, i.e. the questionnaire and interviews. Chapter 5, the analysis, presents the results of the study, both qualitative and quantitative; this is followed by Chapter 6, the discussion, in which the findings of the study are discussed. Chapter 7, the conclusion, summarises the findings of the thesis and outlines the recommendations and implication for theory and practice and concludes with the limitations of the study.

## **CHAPTER 2 CONTEXT OF THE STUDY**

### **2.1. Contextual background to the study**

This section will present a detailed overview of the background to the current study, which sheds light on why teachers hold certain beliefs or attitudes about the OOPT and about Saudi students. As little information is known outside Saudi Arabia about the Saudi Educational system, it is essential to understand the behaviour and the beliefs of the participants under investigation.

The Kingdom of Saudi Arabia (KSA) was first established in 1932 by King Abdulaziz Al Saud. It is governed by a hereditary monarchy, whose King acts both head of state and prime minister (Alhawsawi, 2013). The country has a population of around 27 million people and its official language is Arabic. The country's official religion is Islam and the government adopts Islamic law, which influences many aspects of everyday life, including the education system. The discovery of oil in Saudi Arabia took place around the late 1930s, however it was not until the 1970s that the oil industry boomed and transformed the country's economy (Alkharashi, 2012). The country has since gathered enormous wealth, which has in turn led to improved lifestyle as well as government spending in sectors such as infrastructure, education and healthcare (ibid.). Many expatriates from various different countries were hired to address the shortage in citizen manpower. Over 30% of the manpower is foreign, the majority of which come from Asian countries (e.g. Pakistan, India, Philippines and Indonesia) or other Middle Eastern countries (e.g. Lebanon, Yemen and Egypt). However, while a large portion of Saudi citizens have benefitted from the wealth generated as a result of the oil industry, there are still numerous cases of families living in poverty.

Millions of Saudis struggle on the fringes of one of the world's most powerful economies, where jobs and welfare programmes have failed to keep pace with a population that has soared from 6 million in 1970 to 28 million today.

(The Guardian, 2013)

Saudi Arabia invests a large portion of its budget in education and as a result the literacy rate in the kingdom has risen from 35% to 96% in a period of forty years (Royal Embassy of Saudi Arabia, 2012). Most Saudi citizens wishing to study abroad are also eligible to receive

funding to cover tuition and maintenance for the duration of their studies. In 2006, King Abdullah established the *King Abdullah Scholarship Programme*, which has provided scholarships for thousands of students to attend schools and universities around the world (ibid.). This incentive has encouraged both male and female students to continue their education as well as making HE more accessible to students from disadvantaged backgrounds.

## **2.2. The Saudi educational system**

The Saudi educational system has undergone much transformation in recent years particularly when compared to the way education was confined to the study of Islam in mosques and Quranic schools in the early 20<sup>th</sup> century and earlier years (Al-Liheibi, 2008; Alsharif, 2011). The present educational system can be attributed to the Directorate of Education, founded by King Abdul-aziz, which established a formal educational system in 1925 (Alsharif, 2011). In 1932, the country was unified and became known as the Kingdom of Saudi Arabia (KSA). Since its establishment, it has been the responsibility of the Directorate of Education to open new schools around the kingdom. In the 1930s, the first public schools were open with exclusively male students (Alsharif, 2011; Wiseman, 2010) and the first public schools for girls were opened as late as 1960 (Al-Zarah, 2008). Girls and boys attended different schools, which are segregated in order to comply with the Saudi interpretation of Islamic teachings. At first, the education of girls and women was strongly opposed in some regions by people who did not see a purpose of non-religious education for girls since there were not going to seek employment. However, this view changed quite dramatically after a few years, with a majority in favour of girls' education. (Almutairi, 2008). Saudi education still provides both schooling and HE in entirely segregated environments (Al-Zarah, 2008). While the segregation of students may be linked to beliefs of Islam, cultural and social traditional values also play a role (Wiseman, 2010).

There are a number of different types of schools in Saudi Arabia (e.g. private schools, government schools, international schools), and while these may differ in their curriculum and approach towards teaching, the government mandates that students across all schools adopt the same examination system, with the exception of primary schools since there is no government exam. This ensures that, at least to a certain extent, student outcomes may be



comparable across the country. Progression to the next year of study or from intermediate to secondary school is based upon students' results in their examinations. In some instances, re-sit exams are permitted but, in the event, where a student fails the examination twice, they will need to repeat the school year. As Al-Sadan (2000) points out 'the regulations and procedures of assessment in Saudi Arabian schools omit any reference to individual or group work' (p. 154). This has implications for the way exam results are perceived. Both teachers and pupils thus focus primarily on passing the exams (ibid.). More recently, the teaching of English as a foreign language, as well as English medium schools (e.g. International schools), have gained popularity. The following sections explore the teaching of English in Saudi schools and HE in greater detail.

### **2.3. The teaching of English in Saudi Arabia: An overview**

The teaching of English is highly valued in the Saudi educational system as most citizens regard competency in the language essential both for the individual as well as for the prosperity of the nation (Al-Seghayer, 2014). The teaching of English however, remains fairly recent as Saudi Arabia decided to introduce English into its educational system in order to facilitate communication within the oil industry as well as to deal with Muslims from all around the world visiting the two Holy Mosques (Meccawy, 2010). According to urRahman and Alhaisoni (2013): 'due to the global demand and being the language of 'science and technology, business and commerce' 'window on the wall' etc. the importance of English language grew rapidly (p. 113)'. Therefore, English has become an established subject in the school curriculum, which is administered by the MOE. In the early 2000s, the Saudi government implemented some reforms in the educational structure in order to meet the increasing professional demand for English. For example, they started introducing the language in public school in the seventh grade (first intermediate) and in 2003/2004 English was taught in sixth grade. In 2010, the MOE decided to introduce English in fourth grade (Alrashidi and Phan, 2015) in order to obtain more "fruitful" outcomes. (urRahman and Alhaisoni, 2013). One reason to commence teaching English in primary schools at age 9 was based on the belief that interference might occur when learning English and Arabic at the same time (Alrabai, 2015; Alrashidi and Phan, 2015). For this reason, English was not introduced to younger children.

The MOE has issued general objectives for the teaching of English as a Foreign Language (EFL) in Saudi schools.

### **The Ministry of Education's objectives for the teaching of English**

- To enable students to acquire basic language skills (listening, speaking, reading and writing).
- To develop students' awareness of the importance of English as a means of international communication.
- To develop students' positive attitudes towards learning English.
- To enable students to acquire the necessary linguistic competence required in various life situations
- To enable students to acquire the necessary linguistic competence required in different professions
- To develop students' awareness about the cultural, economic, religious and social issues of his or her society and prepare him or her to participate in their solutions
- To develop the linguistic competence that enable students, in future, to present and explain Islamic concepts and issues, and to participate in spreading Islam
- To enable students linguistically to benefit from English-speaking nations, that would enhance the concept of international co-operation
- To provide students with the linguistic basis that would enable them to participate in transferring other nations' scientific and technological advances

(urRahman and Alhaisoni, 2013)

English is also adopted as the medium of instruction in most Saudi universities, in particular for the teaching of more scientific courses such as medicine and engineering, while courses in the humanities are taught in Arabic. However, universities require all students, regardless of their specialism, to complete compulsory units in EFL. For example, a student studying towards a degree in history must complete an English unit as part of the degree programme. The English requirement has the aim of enabling students to employ the language to access knowledge, which is not available in Arabic. Two universities, which specialise in science degrees, have also adopted English as the sole language of instruction for all courses offered. These are King Fahad Petroleum and Mineral University (KFPMU) and King Abdullah University of Science and Technology (KAUST). However, as a result of a lack of suitably qualified Saudi teachers (holding MA and PhD degrees), the majority of English lecturers in HE are foreigners, predominantly from Arab countries such as Egypt, Jordan and Lebanon (Alrashidi and Phan, 2015).

### **2.3.1. Public and private schools**

There are some key differences between public and private schools in Saudi Arabia, particularly with regards to how much time is dedicated to the teaching of English. There are two main types of private schools: international schools (instruction in English) and private schools (instruction in Arabic). While international schools teach all subjects in English with the exception of Arabic, which is taught as a separate subject often in the context of Islamic education, other private schools are more similar to government schools, in that most of the subjects are taught in Arabic. In these, schools English is taught as a foreign language, however much emphasis is placed on the subject and students begin studying English at a much younger age.

Private school may be defined as ‘owned and governed by entities that are independent of any government – typically, religious bodies or independent boards of trustees that receive funding primarily from non-public sources’ (Alt and Peter, 2002, p.1). On the other hand, public schools may be described as ‘state and local education agencies and publicly elected or appointed school boards [which receive] funding from local, state, and federal’ (ibid.).

The MOE provides textbooks and health services for students regardless of whether students are in private or public schools (Oyaid, 2009). Private schools are supervised by the MOE personnel and must adhere to its rules. Private schools teach the same curriculum and use the same textbooks for some subjects (i.e. Arabic, Islamic Studies, Social Studies, Mathematics taught in Arabic) that are used in public schools (Badawood, 2003). This means all education policies in the country are subject to government control.

State-funded or government secondary schools offer three years of education for students aged 15-18. During the first year students share a common curriculum, while in the final two years students are split up into a “science” path or “humanities” stream. Students who score 60 percent or higher in all first-year subjects have the option of choosing between a science or a humanities specialism, however if they score lower than 60 percent, they must opt for the latter. The curriculum generally includes the following subjects: Arabic, biology, physics, chemistry, English, geography, history, home economics (for girls), mathematics, physical education (for boys) and religious studies.

In public schools, students are assessed throughout the year and are graded through a variety of different tasks and tests, which include the following: during term assessments (i.e. attendance, class participation, homework, midterm quiz and research project) 50% and final exam 50%. Students pass the unit if they achieve at least 50% of the total mark, with 20/50 (40% or higher on the final exam) (Alrashidi and Phan, 2015).

With regards to the education of the children of expatriates, employers generally cover all or most of the basic educational expenses as part of the employment contract. Arabic speakers who are not Saudi citizens can choose to enrol their children in a private school, where the curriculum is almost taught in Arabic, or in an international school. There are also two types of international schools: the first type is run by embassies to serve the needs of their citizens. For example, there are schools for Indian, Filipino, Indonesian, American, and British children, however nationals of other countries, including Saudi nationals, can attend these schools. The curriculum taught in these schools reflects that of the child's home country. Parents of children attending these schools often are interested in courses, which are accredited in their home country in order to facilitate transition upon their return home. The second type of international school is privately run and it admits students from all over the world. In this type of school, English is the language of instruction, and the curriculum generally conforms to the one adopted either in British or American schools.

The majority of teachers in public schools are Saudi citizens, including those teaching English. The minimum qualification is a Bachelor's degree but they are not required to have previous experience or training (Alfahadi, 2014). Most of the teachers graduated with a degree in education or in the arts from local Saudi universities and colleges, which offer four-year BA degrees in teaching English literature and linguistics or translation. There is also an increase in demand for private schools, which provide the same curriculum as public schools but offer additional courses in English language. English teaching in private and international schools starts from an early age, as early as kindergarten, whereas in public schools it starts later, in fourth grade. Students who attend private schools hence receive an average 12 years of English instruction although the exact number of hours and type of instruction varies widely depending on the type of school and its philosophy on English education (Deraney, 2015). Even in public schools, as students move up to the intermediate and secondary stages

the number of English classes increases from 2 to 4 classes per week. Each class lasts 45 minutes (Alrashidi and Phan, 2015). This emphasises the importance of English also from the perspective of parents, who are ready to invest in their children's education from an early age.

Students' schooling also has implications for preparedness for university. According to Al-Seghayer (2011), first year students generally enter university with very diverse proficiency levels, which are linked to their school experience. Students who attended private schools had about twelve years of English, whereas students from public schools had around nine years of English instruction. This, in turn, may mean that students who attended public schools may be required to take a larger number of English language modules in their preparatory year. English language courses are further discussed in section 2.5.1.

### **2.3.2. Overview of Higher Education in Saudi Arabia**

The rapid increase in population and demands of the job market have led to vast transformations in HE. After the establishment of King Saud University in 1957, there were other six universities, which opened over the following 20 years:

- Islamic University was established in 1961
  - King Fahd University for Petroleum and Minerals was established in 1963
  - King Abdul-Aziz University was established in 1967
  - Um Al-Qura University was established in 1967
  - Imam Muhammad Bin Saud Islamic University was established in 1974
  - King Faisal University was established in 1975
- (Alamri, 2011, p.88)

Now, there are 28 public universities and 30 private higher-education institutions in Saudi Arabia (Hamdan, 2015). Saudi Higher Education is quite young; about 65% of government universities were established in the last two decades. The total number of students enrolling in government universities was 1,165,091 in 2014 (Ministry of Higher Education, 2014).

The rapid increase in population and demand for HE has also led to recruitment of non-Saudi nationals. However, while this presents attractive employment opportunities for some, Saudi faculty members generally receive higher salaries compared to their expatriate colleagues and there are also differences with regards to the incentives offered to pursue publications

(Alamri, 2011, p. 90). There are however exceptions; for instance, nationals of Western countries, such as the USA, U.K. or Australia, often native speakers of English, are generally offered more generous pay compared to Saudi nationals. The rationale is to encourage highly qualified individuals to join Saudi universities, as their expertise can improve the quality of education students receive. This also emphasises the importance that English has gained in Saudi education throughout the years.

As a result of an increased demand for competency in the English language, HE has also undergone a number of changes, including placing greater emphasis on developing university students' proficiency in English. One way that universities have addressed this, is to require all university students across the country to either have a high score in a proficiency test (such as the IELTS or TOEFL) upon entry or complete the required English language courses offered in the mandatory preparatory year.

#### **2.4. King Abdulaziz University:**

King Abdulaziz University (KAU) was established in 1967 as a government university with the aim of promoting HE in the western region of Saudi Arabia. The university opened to students in 1968 with a preparation study programme. At the time there were only 68 male students and 30 female students. The first college to be inaugurated was the college of Economics and Management followed by the college of Arts and Human Sciences, which opened the following year. The university rapidly gained popularity and now occupied a distinguished place among the other universities in the Kingdom. KAU is recognised as a prestigious university in Saudi Arabia as a result of the unique professional disciplines and specializations offered. Some of these include Marine Sciences, Meteorology, Earth Sciences, Nuclear Engineering, Mining, Medical Engineering and Aeronautical Engineering ([www.unipage.net](http://www.unipage.net)).

KAU has two campuses, one for male students and one for female students, in order to comply with Saudi Islamic regulations. Both campuses are well resourced, with large libraries, athletic facilities and up to date technology. The university is well known both

nationally and internationally and offers competitive degree programmes that prepare students well for the job market (<https://eli.kau.edu.sa>).

The university was ranked fourth in the 2018 edition of the QS Arab Region University ([www.topuniversities.com](http://www.topuniversities.com)). With regards to admission requirements, generally, any student who has successfully completed secondary school may be admitted to the university, however in order to be admitted onto a science programme, they will need to have higher high school grades. Admission requirements therefore vary according to the degree programme.

## **2.5. The preparatory year programme (PYP) at KAU**

The first year in Saudi universities is referred to as foundation or PYP. For most Saudi universities, this is a fairly recent practice (Al-Shehri, 2017). PYPs provide a ‘middle’ stage between secondary and tertiary education (Al-Maliki, 2013). PYPs were designed to maintain and promote the quality of academic programmes, increase rates of graduation and address the needs of the job market. The current trend in HE in Saudi Arabia is to enhance incoming students’ language, technical and cognitive skills through PYPs (Al-Aqeeli, 2014).

The concept of PYPs in Saudi universities is similar to foundation programmes, in terms of goals, but the content of these programmes is often different. Foundation year programmes in other countries, for instance in the UK, are offered to students who do not meet the conventional entry requirements. These programmes thus focus on preparing students for their specialisation as well as developing their English language skills; they are therefore specifically designed to enable students to study effectively (Arum and Roksa, 2011). PYPs in Saudi universities, on the other hand, focus primarily on developing students’ English language skills and providing opportunities for them to engage with academic English. The content is often related or similar to what they studied in high school. The central aim of setting up a PYP in Saudi Arabia was to equip pre-university students with skills needed for the ‘new century’, bringing them ‘up to par with citizens of the knowledge-based world and enabling them to communicate effectively with the rest of the world’ (Rauf, 2015, p. 185). While the most significant aspect of these programmes is the teaching of English although it should be noted that PYPs also include other general education modules, which students are required to take. The required general education (GE) courses may differ slightly depending

upon the programme of study (e.g. whether students major in the arts or sciences).

However, previous studies on PYPs have suggested that both students and teachers felt that English, as the medium of instruction, was a major challenge of these programmes (Al-Adwani and Al-Abdulkareem, 2013). Students' English language skills upon entry are often weak, particularly if they attended public schools, and, as a result, many students prefer universities or specializations that do not require completing preparatory year programmes (Al-Shehri, 2017). According to Al-Maliki (2013), PYPs therefore, play a central 'middle stage' between public and tertiary education. Through these programmes students not only improve their English but also develop transferable skills such as communication and research (ibid., cited in Al-Shehri, 2017).

### **2.5.1. English courses within the PYP at KAU**

As mentioned above, the teaching of English is the most significant aspect of PYPs and student success in the compulsory English language courses is required for admission onto any degree programme.

The English Language Institute (ELI) at KAU was first accredited by the Commission on English Language Programme Accreditation for the period April 2013 through April 2018 and newly accredited for a further 10 years period. The programme therefore, agrees to uphold the Commission of English (CEA) Standards for English Language Programmes and Institutions.

*(<https://eli.kau.edu.sa>).*

The CEA is recognised by the U.S. Secretary of Education as a national accrediting agency for English language programs and institutions in the U.S. and also conducts accreditation activities internationally. This accreditation was regarded as an outstanding achievement for the ELI when it became the first CEA accredited institute to teach an English language program to Preparatory Year students at a Saudi university.

It should be noted that English language programmes in Saudi universities are often influenced by policies and practices from both American and British higher education. This



is also evident in the usage of terms such as preparatory and foundation year, which in Saudi Arabia are often employed interchangeably.

The ELI's mission is to help students achieve an intermediate level of English proficiency equivalent to the Common European Framework Reference for Languages (CEFR) B1 Threshold or Intermediate Level within one year. The short-term mission of the ELI is to ensure that students are fully equipped with the necessary level of English proficiency that is needed for their undergraduate studies in various KAU colleges, especially among students whose majors will be taught in English (e.g. Medicine and English Literature). On the other hand, the long-term mission is to prepare students to compete in their future careers, as English is significantly valued in any job market. (<https://eli.kau.edu.sa>).

According to Moskovsky and Alrabai (2009), 84.9% of Saudis tend to learn English to get better-paid jobs, and proficiency in the English language has become a prestigious skill in the Saudi community. However, a limited number of students who apply for the English Literature degree get accepted. Every year the number of students admitted amounts to around 20 to 25 students per semester. While the English Literature Department's entry requirements are high, low student numbers could also indicate that the better students choose a science specialism. The Department requires students to score a 75 or above in their English modules, which sets a difficult standard to achieve.

It should be acknowledged that while students enter university with some previous knowledge of English, teaching methodologies differ considerably between the school and HE sectors. In schools, students are generally taught English through grammar-translation based approaches, while PYPs adopt communicative language teaching. This means students are assessed on all four skills: reading, writing, speaking and listening. The programmes are generally multi-level (there are different ability levels e.g. level 1, 2, 3, 4) and include a variety of assessment tools (Rauf, 2015) meaning that students are assessed throughout the course rather than sitting one exam at the end of the module. While some students may welcome the more interactive approach towards learning English, others find it very challenging.

PYPs aim to develop students' English language skills within a short timeframe, however, as Al-Shehri (2017) points out, it would appear that PYPs are not able to effectively improve students' language skills in just one academic year (see Al-Asmari, 2013; Al-Otaibi, 2015; Al-Juhani, 2012). Al-Shehri (2017) stresses that it was clear that for some students the English language itself had become the main obstacle. This particularly affected students who were admitted onto the PYP with poor language proficiency and limited language skills. Failure to meet the minimum requirements in a PYP may result in students not being admitted onto the programme of their choice or, in some cases, disqualify them from being able to pursue their studies in any university programme (Al-Shehri, 2017, p.436).

While at KAU the English Language component of the preparatory year is not taught for specific purposes, this is not the case across all universities in Saudi Arabia. Some universities offer different English Language routes. At Qassim University for example, the PYP for students in the sciences is divided into two semesters. In Level 1, the first semester, students study general subjects like English for General Purposes (EGP) as well as Statistics, Physics, Computer Science Course (CSC) and Thinking Skills, however, depending upon their performance in Level 1, students are then given the option to either study medicine or science. Students are then taught English for Specific Purposes (ESP) and the content focuses on their subject specialism (Hussain et al., 2017).

The English courses offered at KAU differ according to proficiency levels and include a range of assessments, which provide teachers with several opportunities to assess students' progress throughout the course. A syllabus for each of the four ELI modules is available on the ELI website. The documents mainly outline the course objectives, required materials and assessment (see level 1 syllabus in appendix 1). The following section describes the current assessment procedures in place at KAU.

## **2.6. Current admission and assessment procedures in the ELI**

The ELI offers four English Language modules (level 1-4), which are all credit-bearing courses, and thus students' grades contribute to their overall GPA, with the exception of level 1. Level 1 is the only level, which is not credit-bearing and thus students are just required to pass this course in order to be able to progress to level 2. The figure below illustrates the credits, levels and weekly contact hours for each module.

**Figure 1. ELI courses – Summary of levels and credits awarded**

ELI COURSE	COURSE LEVEL	CEFR LEVEL	CREDITS	WEEKLY CONTACT HOURS
ELI 101	Beginner	A1	0	18
ELI 102	Elementary	A2	2	18
ELI 103	Pre-Intermediate	B1	2	18
ELI 104	Intermediate	B1+	2	18

[http://eli.kau.edu.sa/Default.aspx?Site\\_ID=126andLng=EN](http://eli.kau.edu.sa/Default.aspx?Site_ID=126andLng=EN)

Placement into ELI courses is initially determined by the scores students achieve in the OOPT, which is a university requirement for admission onto any degree programme. In order to progress to a higher-level course (e.g. from ELI 101 to ELI 102), students are required to pass the ELI assessments.

### 2.6.1. The OOPT at KAU

The ELI delivers its English language programme to approximately 15,000 full-time male and female students annually, across the three campuses. Some students are likely to be exempted from the English preparatory year, if they score satisfactorily on the General IELTS (4.5) or TOEFL IBT (45 or above). If they do not qualify for an exemption, they are required to sit the OOPT placement test. This is a mandatory test, and it is offered only once in order to place students in the appropriate level.

As specified on the ELI website, KAU requires that students pass Level 4 (ELI 104, CEFR B1+) in order to secure entry into their degree programme. As part of the registration process for the preparatory year at KAU, all students are required to attend a presentation on the OOPT, which provides an overview of what the test will look like on the computer screen and instructions on how to navigate (e.g. which buttons students need to press to move on to the next page or submit their test). The presentation is scheduled on the same day that students take the OOPT and focuses primarily on the technical side of the exam. If students need assistance with technical aspects while taking the test, a few technicians are available to resolve such issues.

Student registration takes place during the summer and the ELI orientation, which provides students with an overview of the English language requirements and the OOPT also occurs during the two-weeks registration period. The presentation does not provide students with advice with regards to the content of the exam, however all students are informed that the OOPT is a university requirement during the application process. If a student does not show up for the OOPT, she will automatically be placed in the lowest level (ELI 101), regardless of her proficiency in the language. A student may be moved to a higher or lower level upon the recommendation of the teacher, however in exceptional cases, where a large number of students complain about their test results, these students may be allowed to re-take the OOPT.

#### **2.6.1.1. Content and structure of the OOPT**

The OOPT was developed by Oxford University Press (OUP) with the aim of evaluating learners' proficiency in their linguistic knowledge and their listening ability. It is a tool that is meant to be valid and reliable in measuring students' language knowledge and how that knowledge is used in communication (Oxford University press, 2015).

The OOPT is a computer-adaptive exam that is automatically marked, meaning that it adapts to the ability level of each participant. It estimates the ability of each test taker after answering a question and then selects the most suitable item to present next, according to their level. If a test taker answers the question correctly, a harder question will be provided and vice versa. Students are placed in labs that are equipped with headphones to use for the exam. The test takes around 60 minutes to complete. According to Purpura (2007, p.2), the OOPT measures the following:

- **Section One: The Use of English (~30 questions)**

Section One measures students' knowledge of the grammatical and pragmatic aspects underlying the English language. It consists of four language knowledge exercises. The first exercise measures grammatical forms. The second task measures semantic meaning. The third task measures grammatical form in addition to meaning in cloze format. The last exercise measures students' knowledge of pragmatic meaning encoded in interaction.

- **Section Two: The Listening Section (~15 questions)**

Section Two measures students' listening ability. It consists of three tasks that measure students' ability to understand both the literal meaning and the implied meaning within the text or beyond the parameters of a listening text. The tasks are presented as short dialogues, long dialogues and monologues.

The questions in the listening section are in many formats: the first one presents the student

with a number of short dialogues; each dialogue is followed by a single four-option multiple choice question. The second format, on the other hand, presents a long dialogue, while the third format presents a monologue. Students are allowed to listen to each recording twice if they wish.

Unlike traditional tests, the OOPT score ‘is a measure of how difficult the hardest items were the student could succeed on... it is a judgment rather than a count of anything’ (Pollitt, 2009, p.11). The OOPT presents questions in a rather authentic communicative context, which aims to test students’ ability to understand how language is actually used and the meaning, which can be inferred from conversations. The sample questions below, taken from the OOPT website, illustrate how ‘Use of English’ questions attempt to incorporate an authentic cultural context.

**Figure 2. Sample OOPT questions – Use of English**

**Use of English \* Section 1: Approximately 30 Questions**

**Read the dialogue. Then, select the correct answer from the options below.**



**Use of English \* Section 2: Approximately 30 Questions**

**Read the conversation. Then, select the correct answer from the options below.**



## 2. What does the boy mean?

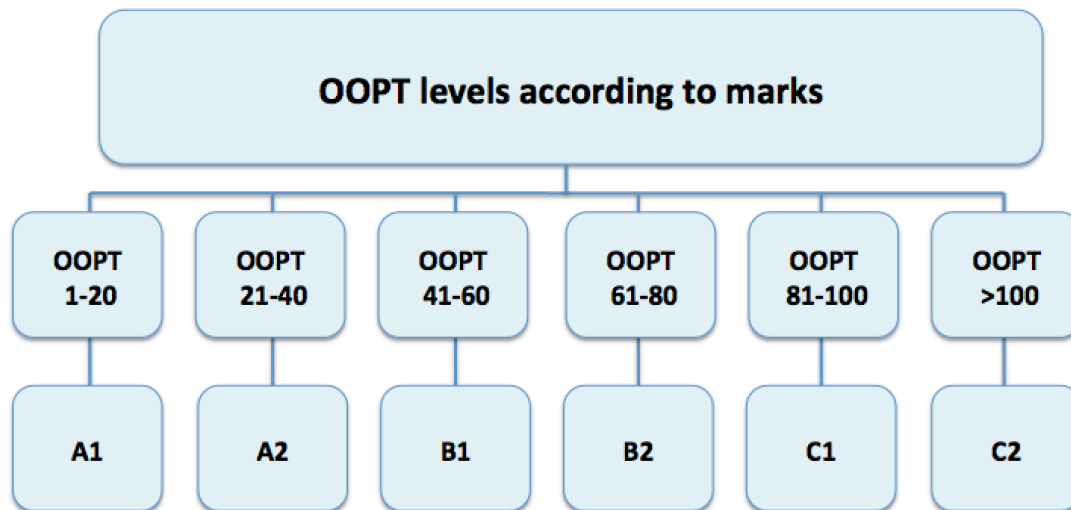
- ☐ A I can eat an apple.
- ☐ B I have got an apple.
- ☐ C I would like an apple.

As can be observed from the sample above, the questions are multiple choice and students are provided with a brief interaction, from which they have to infer the correct meaning.

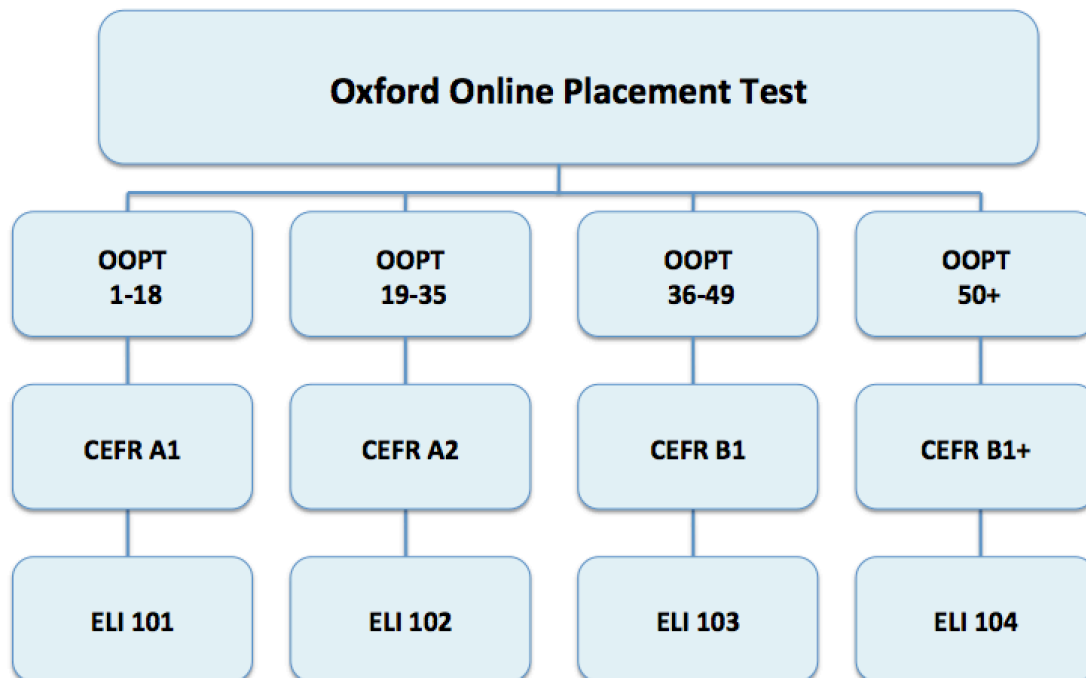
### 2.6.1.2. OOPT levels of proficiency vs. ELI levels

A total of 120 points are possible on the OOPT, and scores are equated to one of the 6 levels of the common European Framework of Reference (A1-C2). However, at KAU, only four levels are offered, hence the OOPT levels determined by the marks are different from those adopted by the university's English language programme. Figure 3, illustrates the difference between the OOPT levels, which are based on the CEFR, and the four levels offered by the university.

**Figure 3. OOPT levels and ELI levels according to OOPT marks**



(Adapted from <http://www.oxfordonlineenglishtesting.com>)



(Adapted from <http://eli.kau.edu.sa>)

Pollitt, a member of the UK OFQUAL Advisory Committee, claims that:

The OOPT has been designed in order to serve one primary purpose – the accurate placement of

students learning English into a class or a course of teaching that is appropriate to their current levels. Anyone using it for this purpose can be confident that it is valid for that use.

(Pollitt, 2009, p.13)

KAU decided to adopt this placement test because it considers the OOPT to be a quick and inexpensive way to measure a student's general English language ability. This makes it preferable to the IELTS and TOEFL, which KAU considers to be time-consuming and demanding, as preparation is required prior to the test. Furthermore, the IELTS and TOEFL were not designed as placement tests.

The ELI courses at KAU are taught using the *Cambridge University Press English Unlimited Special Edition* textbook series. Each ELI module runs for a period of seven weeks per semester, in which students have 18 contact hours a week. These are intense language courses, which focus on developing the four skills of listening, speaking, reading and writing as well as developing students' knowledge and application of grammar and vocabulary. The program aims to rapidly develop students' competencies in English throughout the preparatory year.

Yet, while for those students entering with at least an intermediate proficiency in English the 18 contact hours a week may be enough, this is unlikely to be sufficient for those students entering into level 1. A study carried out at KAU researched students' OOPT results and found that the majority of students who took the placement test were placed in level 1 and 2.

The results of the placement test were really alarming. Most of the students lack the basics of English language in the four English language skills. More than 87% of students fall in level one and level two according to the results of the placement test. This is due to the huge gap between the school education and the university education in KSA.

(Mahmoud, 2014, p.337)

As Mahmoud (2014) points out, it is possible that one of the reasons why students underperform on the OOPT is that students are being tested on their understanding of the 'use of English' and 'listening', two skills in which they will have had little or no prior practice in school.



### 2.6.2. ELI achievement tests

According to the KAU website, the ELI assessed students across the four English language modules throughout the course through *continuous assessment* as well as more formally through a mid-term and final exam. There are five different types of assessments named as follows: 1. *Formative Continuous Assessment*, 2. *Speaking Examination*, 3. *Writing Examination* 4. *Computer-based Mid-Module Examination*, 5. *Computer-based End of Module Examination*. The figure 4 below, illustrates the weightings for each of the assessments as well as giving a description of the format.

**Figure 4. Assessment in ELI modules**

Method	Format	Weight
<b>Formative Continuous Assessment</b>	<ul style="list-style-type: none"><li>• Writing Tasks 10 %</li><li>• Grammar and Vocabulary Use 10 %</li></ul>	20%
<b>One Speaking Examination</b>	Range from basic interviews with leading questions (lower levels) to extended turns and discussions (higher levels). Time allowed: 3-5 minutes.	10%
<b>One Writing Examination</b>	Range from constructing simple sentences and short paragraphs (beginner level) to more comprehensive, cohesive paragraphs at Elementary and Pre-Intermediate level, to writing short, coherent essays (Intermediate level). Time allowed: 40 minutes.	10%
<b>Computer-based Mid-Module Examination</b>	Multiple choice questions with focus on reading and listening comprehension, and vocabulary and grammar use from units covered in the first three weeks of the module. Time allowed: 90 minutes.	20%
<b>Computer-based End-of-Module Examination</b>	Multiple choice questions with focus on reading and listening comprehension, and vocabulary and grammar use from units covered in the entire module. Time allowed: 105 minutes.	40%
<b>Total</b>		<b>100%</b>

[http://eli.kau.edu.sa/Default.aspx?Site\\_ID=126andLng=EN](http://eli.kau.edu.sa/Default.aspx?Site_ID=126andLng=EN)

All the assessments are drawn from the material covered in class, which is based upon the Cambridge textbooks adopted throughout the four levels. The *Formative Continuous Assessment* includes writing tasks and quizzes on grammar and vocabulary use, which are administered throughout the course and contribute to the students' final grade in the module. The Computer-based end-of-year and mid-module exams are much broader and cover a range

of skills, including reading, listening, vocabulary use and grammar.

Progression to a higher ELI level is based upon achieving a satisfactory result in the assessments. Students must achieve a minimum overall grade of 60% in order to pass the module and move up to the next level. The ELI website emphasises that ‘great care has been taken to ensure that the ELI assessment instruments measure achievement of each level’s student learning outcomes (SLOs)’.

The SLOs specified in the ELI course syllabi stem from SLOs originally formulated in close consultation with OUP. These are ‘currently being adapted in-line with the newly adopted (2015-2016) Cambridge University Press *English Unlimited Special Edition* instructional materials’ (ELI 2017/18 Student Handbook, p. 35).

## **2.7. The Cambridge Unlimited Special Edition Materials**

The ELI chose to adopt the *Cambridge Unlimited Special Edition* in the 2016-17 academic year and these materials are currently in use for all the four levels taught at KAU. In previous years, the ELI employed Oxford’s *New Headway Plus Special Edition*, which had been used for a number of years. The introduction of the Cambridge materials meant that SLO, as well as assessments, had to match the new curriculum. It should be noted that for both the previously employed Oxford materials as well as the recently introduced Cambridge ones, the *Special Edition* was adopted by the ELI. Unlike the standard editions, special editions are specifically designed for speakers of Arabic and are therefore regarded as more culturally appropriate for Saudi students. The Cambridge University Press website describes the *Cambridge Unlimited Special Edition* as follows:

English Unlimited Special Edition is a CEFR goals-based course for adults. Centred on purposeful, real-life objectives, it prepares learners to use English independently for global communication. Through universally inspiring topics and activities, and with a special focus on the learning needs of Arabic speakers, this truly international course helps learners become more sensitive, more effective communicators. Teaching natural, dependable language and using authentic audio from the start, it not only brings real life into the classroom, but gives learners the skills, strategies and confidence they need to communicate confidently outside it.

(<http://www.cambridge.org/ps/cambridgeenglish/catalog/adult-courses/english-unlimited-2>)

The *Cambridge Unlimited Special Edition* has four different course-books available for CEFR levels A1, A2, B1 and B1+, which are the four books adopted for the ELI English language courses. Cambridge *does* offer two other coursebook packages (one for CEFR B1 and one for CEFR C1), however these are only available in the standard edition (<http://www.cambridge.org/ps/cambridgeenglish/catalog/adult-courses/english-unlimited-2/toc> ).

The fact that a special edition is not currently available for levels higher than B1+ may be a contributing factor to the ELI's decision to offer four courses rather than six.

One of the issues emphasised in the literature, however, relates to the way the CEFR level descriptors have been interpreted by teachers. Fulcher (2004) argues that the impact of the CEFR on the adequacy of its underlying construct needs to be further explored:

For teachers, the main danger is that they are beginning to believe that the scales in the CEFR represent an acquisitional hierarchy, rather than a common perception. They begin to believe the language of the descriptors actually relates to the sequence of how and what learners learn.

(Fulcher, 2004, p. 260)

Clearly, there will be students who are placed in higher CEFR levels, who may be weaker than others in some areas and yet may be performing above their current level in other skills.

## **2.8. Teachers' background**

Teachers in the ELI come from a variety of different backgrounds. The majority are not Saudi Citizens (for the most part from Egypt, Pakistan, Tunisia or India and a few come from English-speaking countries such as the USA, UK or South Africa). There are also a large number of Saudi teachers. With regards to Saudi teachers, they are required to hold a Bachelor's degree in relevant area such as English literature, linguistics or translation. Most foreign teachers, on the other side, hold higher qualifications compared to Saudi citizens; they may hold a postgraduate degree, an English teaching qualification or be more experienced e.g. they may have several years of teaching experience.

Saudi teachers generally hold permanent contracts, while non-Saudi citizens are offered yearly contracts, which can be renewed. Teachers are not required to have previous knowledge of testing, nor does the university hire additional staff with this expertise.

## **2.9. Students' background**

The majority of university students in public universities in Saudi Arabia are Saudi nationals as non-nationals are not entitled to apply for a place in a public university (except in rare cases e.g. where the student has applied for a scholarship to study Islamic Studies or Arabic) or if their mothers are Saudi citizens. Students applying for both degrees in the Arts or Sciences are therefore, generally progressing from secondary schools and are between the ages of 18 and 20. Students joining the ELI may have attended either public or private secondary schools. In the final two years of high school, students need to specialise in the Arts or in the Sciences. This affects their choices when studying at university as those who have specialised in the Sciences are given the choice to carry on studying towards a Science degree or major in the Arts, while those who have specialised in the Arts are not allowed to major in the Sciences; they must study towards a degree in the Arts. Furthermore, Jeddah is an attractive city for students, and KAU received applications from students residing in neighbouring villages as well as from other cities in the region.

## **2.10. Summary**

The chapter provided an overview of the educational system in Saudi Arabia, placing particular focus on the teaching of English in the ELI at KAU. The chapter further describes the OOPT and the differences between the ELI CEFR-based proficiency levels and those suggested by the OOPT itself. The context of the study is central to the investigation as it serves to contextualise the complexities behind teachers' beliefs about the placement test, students and the English language modules. Throughout the thesis reference is made to the Saudi context and the curriculum and policies of the ELI described above.

## CHAPTER 3 LITERATURE REVIEW

### 3.1. Introduction

The chapter reviews the literature on test validity by describing how the concept evolved over time and places particular emphasis on the unified view of validity, central to the investigation. It further draws upon the literature on teachers' beliefs about testing, and test takers' characteristics as well as the literature on testing and placement tests placing focus on the influence of contextual factors on test performance.

Messick's (1989) unified view of validity is of particular relevance, as it includes the consequences of test use. In this study, the unified view of validity can be regarded as an "umbrella" under which perceptions, uses and consequences of test use can be interpreted.

Issues related to the use of English language placement tests are also reviewed, as these serve to contextualise the research questions guiding the investigation. The literature also reviews relevant studies carried out in Saudi Arabia, which provide a Saudi perspective on some of the issues, such as the relevance of the content of the OOPT within this cultural context. The summary of the chapter discusses areas in which there are gaps in the literature and, therefore, potential for further investigation.

### 3.2. The concept of validity over time

A major question regarding language tests concerns whether a test *does* indeed *test what it is supposed to test*; this, asks a traditional validity question, and is a key concept in language testing. According to Alderson, Clapham and Wall (1995), 'if a test is not valid for the purpose for which it was designed, then the scores do not mean what they are believed to mean' (p. 170). To answer the previous question, we need to establish a purpose and a valid usage of a test and follow up with questions like: *How do we know if a test is valid?* or *for what it is valid?* (p.170).

Some of the first conceptualisations of validity emerged as early as the mid 1800s with the introduction of written examinations in schools in the USA and Local Examinations for

schools in England (Newton and Shaw, 2014, p.17). By the end of the 19<sup>th</sup> century there was a strong belief in structured assessments, however many began to question whether the results of these examinations were accurate, which led to the development of tests with less subjectivity, such as those with true or false items and multiple choice. This led to the development of standardised testing (ibid.). In 1921, the American National Association of Directors of Educational Research published a document to seek a consensus on the meaning and terms and procedures adopted in measurement movement. In addition to defining key terms, the document also defined validity as ‘the degree to which a test measures what it is supposed to measure [...] this provided the foundation for all subsequent thinking on validity’ (ibid., p. 18).

Towards the beginning of the 1950s, a committee of the APA, chaired by Lee Cronbach, developed the first *Standards* document (published in 1952) which provided guidance on the information that test producers should provide on tests (ibid., p. 19). The document included a section on validity, which classified validity into ‘types’ and ‘aspects’. A few years prior to this publication, Greene et al. (1943) had distinguished between curricular validity and statistical validity. Similarly, Cronbach (1949) made a distinction between logical validity and empirical validity. The APA 1952 draft distinguished four types of validity, namely: content, predictive, status and congruent. In the 1954 final publication, these were revised to: content, predictive, concurrent (previously status) and construct (previously congruent). The concept of **construct validity** had been first proposed by Paul Meehl and Robert Challman. This was later modified by the committee and elaborated into construct validity in psychological tests (Cronbach and Meehl, 1955).

The *Standards* was revised in 1966 with the four types of validity being reduced to three: **content, criterion-related and construct** and it was stressed that the three types should not be seen as mutually exclusive (Newton and Shaw, 2014, p.20). Later, Cronbach emphasised the importance of constructs to all educational and psychological testing (Cronbach, 1971). Cronbach (1971) made a significant shift in the validity theory; the emphasis moved from investigating the validity of a specific instrument to the interpretations of the measurement outcomes. In another words, ‘one does not validate a test, but an interpretation of data arising from a specific procedure’ (p. 447). Cronbach (1971) stressed that there is no such thing as

test validation, but the interpretation from test scores that has to be validated instead. This view has remained constant and many researchers agreed that the focus on the validation process should rely on the interpretation and the usage of test outcomes rather than focusing on validating the test itself.

To get a better understanding of validity in terms of using test results, we have to consider several types of validity, for example: **internal validity**, **face validity**, **content validity**, **response validity**, **external validity** and **criterion validity**. According to Alderson, Clapham and Wall (1995), '**internal validity** relates to studies of the perceived content of the test and its perceived effect' (p.171). Ingram (1977, p. 18) pointed out that **face validity** refers to the test's 'surface credibility or public acceptability' (in Alderson, Clapham and Wall, 1995, p. 172). In this type of validity, students, administrators or non-expert people can comment on the value of the test whose judgment is not 'expert' (p.172). These non-expert users can give comments and judgments about a certain test as a whole.

The third type of validity, **Content validity**, can be defined as 'the *representativeness* or *sampling adequacy* of the content- the substance, the matter, the topics- of a measuring instrument' (Kerlinger, 1973, p. 458, cited in Alderson, Clapham and Wall, 1995, p. 173). Unlike face validity, content validity requires judgment of 'expert'; that is 'people whose judgment one is prepared to trust, even if it disagrees with one's own' (Alderson, Clapham and Wall, 1995, p. 173).

The period between 1974 and 1999 was described as the 'Messick years' as a result of the dominance of his view on validity. Informed by Harold Guilliksen, Jane Loevinger and with the support of Robert Guion, he appeared to have been able to persuade most measurement professionals around the view that validity should be understood as construct validity. Messick later became more interested in how test consequences informed evaluation of test score meaning. He focused on the positive and negative social values associated with consequences of testing, including the misuse of tests (Newton and Shaw, 2014, p.22). The following section describes Messick's unified view of validity, which offers a useful perspective under which to consider the findings.

### 3.3. Messick's unified view of validity

The theoretical concept of validity currently emphasised by psychological testing researchers draws upon the work of Messick (1989), who defined validity as a unitary concept. Messick (1989, p. 18) added a further component to the process of establishing validity:

For a fully unified view of validity, it must also be recognized that the appropriateness, meaningfulness, and usefulness of score-based inferences depends as well on the social consequences of the testing. Therefore, social values and social consequences cannot be ignored in considerations of validity.

Messick's unitary view of validity has become a helpful foundation for language testing researchers, who might consider issues of the interpretation of test result, as well as the consequences of test use (Bachman, 2000, p. 21). This view of 'score interpretations and uses necessarily raise the issue of test consequences' (Alderson and Banerjee, 2002, p. 79).

Messick (1996) defined validity as an, 'overall evaluative judgment of the degree to which empirical evidence and theoretical rationales support the **adequacy** and **appropriateness** of **interpretations** and **actions** based on test scores or other modes of assessment' (emphasis in the original, p. 6).

Messick's view has gained more acceptance among researchers; Fulcher and Davidson (2007) have noted that 'Messick's way of looking at validity has become the accepted paradigm in psychological, educational and language testing' (p. 14). Language testers have come to accept rephrasing the question 'what does our test measure?' or 'does this test measure what it is supposed to measure?' to 'what is the evidence that supports particular interpretations and uses of scores on this test?' (Alderson and Banerjee, 2002, p. 79). Messick's unified model of validity is particularly relevant for language testing and was endorsed by the *Standards for Educational and Psychological Testing* (AERA, APA, and NCME 1999) and 'the consequences of testing were identified in the Standards as one of five sources of validity evidence' (Hubley and Zumbo, 2011, p. 220). In Messick's unified validity theory, it is acknowledged that cultural or social norms and values can change, and this can affect both theory and how constructs are viewed. Similarly, language also changes over time, thus affecting meaning, understanding and the relevance of items in a test (Hubley and Zumbo, 2011). Messick (1989) argued that what needs to be validated are the



*interpretations* linked to test scores, because test scores are not merely a function of test items. Test scores are also a function of the individuals taking the test as well as its context and specific purpose. In his view, validity is an ongoing process to be evaluated whenever a test is used. A test is thus considered valid when there is sufficient theoretical and empirical evidence to allow for interpretation and uses of test scores.

The *Progressive Matrix of Validity* model (Messick, 1989) describes what needs to be taken into consideration when validating inferences both in terms of interpretation of test scores and their uses (see table 1).

**Table 1. Facets of Validity as a Progressive Matrix (Messick, 1989)**

	<b>Test Interpretation</b>	<b>Test Use</b>
Evidential Basis	Construct Validity (CV)	CV+ Relevance / Utility (R/U)
Consequential Basis	CV+ Value Implications (VI)	CV+RU+VI+ Social Consequences

Messick's framework of unified validity distinguishes between two interconnected facets of validity: the source of justification and the function or outcome of testing (Messick, 1993). He argued that where the facet for justification (either an evidential basis or consequential basis) is crossed with the facet for function or outcome (test interpretation or test use), then a four-fold classification is obtained highlighting both meaning and values in both test interpretation and test use (as illustrated above). The four cells in the table correspond to the four interrelated aspects of validity. As the table highlights four different foci in addition to the basic construct validity, Messick described the framework as a *progressive matrix*.

One implication of this progressive-matrix formulation is that both meaning and values, as well as both test interpretation and test use, are intertwined in the validation process. Thus, validity and values are one imperative, not two, and test validation implicates both the science and the ethics of assessment, which is why validity has force as a social value. (Messick, 1993, p. 16)

Messick's unified view of validity is particularly relevant to this study as the investigation explores teachers' beliefs about the OOPT as a placement test as well as their views on the consequences of test use within its context (i.e. a PYP at KAU). It is thus argued that exploring perceptions, uses and consequences of a test can provide information about its validity.

### 3.4. Psychometric properties of tests

Psychometric properties of a test refer to the levels of reliability and validity. Psychometric testing was theorised to build ‘a more scientific way of assessing language’ (Lado, 1961), which drew upon the work of Bloomfield (1933) on linguistic science. Green (2014) stated that psychometric testing was concerned with the technical measurement qualities of tests. While the validity of traditional tests was based on the qualifications and status of the people who designed and built them, psychometric approaches to validity highlighted the need for hard statistical evidence that a test was working effectively. This kind of evidence might include high levels of reliability and similarities between the results obtained from different test scores of the same abilities. Psychometric tests generally avoid looking at individual items or questions, and rather observe the total score of a test taker compared to a representative sample for the same.

### 3.5. Washback: the effects of testing on teaching

The concept of *Washback* was introduced first by Messick in (1989) where the notion of consequences was incorporated with his definition of validity. Messick (1996) defined *Washback* as ‘the extent to which the introduction and use of a test influences language teachers and learners to do things that they would not otherwise do that promote or inhibit language learning’ (p.241).

Messick’s (1989) conception of validity incorporated both the values endorsed by the test, and the impact that the use of the test had on individuals and institutions (Fulcher, 2010, p.277). While other terms, such as ‘backwash’ have previously been employed to refer to ways in which testing may affect teaching materials and classroom management (Hughes, 1989), within applied linguistics and language testing, the term ‘washback’ has become more widely used (Weir, 1990; Alderson and Wall, 1993; Alderson, 2004). There are two types of washback: negative and positive. Hughes (2003) suggested that it can be harmful or beneficial. While positive washback refers to ways in which testing can positively influence teaching and learning, negative washback has been described as ‘a mismatch between the construct definition and the test, or between the content (e.g., material/ abilities being taught) and the test’ (Brown, 2002, p. 3). Alderson and Wall (1993) made a significant contribution

to washback research, in particular in the field of language testing. The authors argue that testing will influence what and how teachers teach, what and how learners learn, the rate and sequence of teaching and learning as well as the attitudes towards the content, method and other aspects of teaching and learning.

However, as studies into the introduction of new tests in education systems would suggest, things do not always work as intended. While there is evidence of possible effects of testing on teaching and learning, there is no simple relationship between the use of a test and its effects (Wall and Alderson, 1993; 1996). Hughes (1994), however, questioned Alderson and Wall's views and argued that there was a need to be more precise about what could be classed as washback. The author defined washback as follows: 'the test's effect on any aspect of teaching and learning' (ibid. p.1) and suggested that this definition could be adequate for most purposes. He also proposed his own way of categorizing different types of washback and distinguished between the effects of washback on 'participants', 'processes' and 'products' of an educational system. Hughes (1994) described 'participants' as the classroom teachers and students, educational administrators, textbook developers and publishers 'all of whose perceptions and attitudes towards their work may be affected by a test' (ibid., p. 2). Furthermore, he defined 'process' as 'any action taken by the participants, which may contribute to the process of learning' (ibid.), such as materials development, syllabus design, changes in teaching methodology, the use of test-taking strategies, etc. Finally, 'product' refers to 'what is learned and the quality of the learning' (ibid.).

Watanabe (2004) pointed out that there is no systematic way in which testing affects learners and that some may be affected more than others. There are, in fact, other factors, which also play a role. As Watanabe (2004) argues, washback may be mediated by factors, such as: the nature of the curriculum, the training background of teachers, and the culture of institutions (in Fulcher, 2010, p. 278). Therefore, researching washback requires careful thought, particularly when investigating within a familiar context (Fulcher, 2010, p. 279).

Wall (1997) argues that 'test washback' needs to be distinguished from 'test impact' as the two concepts describe different ways in which testing may influence teaching and learning. He argued that while *test impact* is concerned with any effect that an examination may have

on individuals, policies and practices in a society or educational system, *test washback* refers specifically to the effects on teaching and learning. The current literature thus gives special emphasis on the importance of washback studies as recent research has shown that it is critical variable, which may also be considered part of test validity (Shepard, 1997). There are currently many studies in applied linguistics, which argue that the consequences of a test may be considered as an integral part of test validity, known as *consequential validity* (Messick, 1996). Since the coining of the term, a number of discussions have taken place among language testers on consequential validity (Kunnan, 2000; Messick, 1996; Shepard, 1997). As a result, most testing experts currently acknowledge that test washback and test impact are complex phenomena and therefore need to be taken into consideration as fairness and ethical factors in designing tests (Hamp-Lyons, 1997).

Alderson et al. (1991) argued that test developers should explore specific areas, such as content, methodology and assessment tools and particularly the extent of the presumed impact of tests on teaching. They pointed out that washback does not only affect individuals, but also influences the educational systems. Hence washback affects more than testing and teaching. Bachman and Palmer (1996) suggested that there are three aspects of testing, that significantly affect test takers: the experience of preparing and taking the test, feedback, and the decisions taken as a result of test scores. During the test preparation period student will practice test techniques, which may be unrelated to the syllabus. Generally, asking test takers regularly about their perceptions of a test will have a positive impact on their performance, as they will have a clear understanding of the task. Nevertheless, there may be unexpected features of the test that may de-motivate them, in spite of being well prepared. Test scores can also greatly affect test-takers, as they will influence the decisions students will need to make in future.

### **3.6. Cultural appropriateness in testing and materials:**

It is acknowledged that test developers face numerous challenges when developing placement tests. On the one hand, they need to design a test which can be used by learners coming from a wide range of cultures and backgrounds without discriminating any groups or individuals, on the other, it has been argued that 'language and culture are inseparable' or

that ‘language and culture are intimately linked’ (see Byram, 1989; Byram and Morgan, 1994) and culture will therefore implicitly feature in any tests. Byram’s view of culture has however been challenged by authors, such as Risager (2006), who disagree with this view and have contested the inseparability of the language and culture. The author proposes a transnational alternative to the national paradigm and from this understanding, language learning may make reference to a culture but not exclusively the nation culture/s of the country or countries where the language is spoken. The divergent theoretical perspectives further add to the complexity of designing English language tests, which are to be used internationally in a wide range of educational contexts with culturally diverse students.

Crozet (2016) similarly emphasised the importance of moving beyond the static transmission of cultural knowledge. Crozet and Liddicoat (1999) emphasised the link between teaching and assessment and argued that cultural content should feature in assessment scales. The authors argue that the inclusion of cultural content can radically transform the very way we have conceived the goals of language teaching and learning. Crozet and Liddicoat (1999) further suggested that a second language is learnt in order to be used and that language use is fundamentally cultural. This view draws upon the work of Kramsch (1993) who argued that language learners need to develop a cultural position, which mediates between their own culture and the target culture. Assessments can reflect this understanding of culture and include examples of cultural mediation rather than facts about the target culture.

Researchers have highlighted the danger of ‘essentialising’, i.e. allocating particular characteristics to (national) cultures. Rather language teaching should provide a context for fostering intercultural competences more generally. The concept of global English is a good example of how language has been disassociated from its target culture(s) thus challenging the concept of a “native speaker” model, which learners should aspire to. Byram and Grundy (2003) argued that learners should have opportunities to be introduced to unfamiliar contents, which can then be compared with the learners’ own culture.

As a result of the complexity behind the relationship between culture and language teaching, the inclusion of culture in EFL textbooks and proficiency tests has been a heated discussion. In particular, because English is widely taught and employed as a lingua franca across the

globe, some argue that textbooks should present the target culture and its values, while others feel that materials should draw upon the students' own culture. It has been suggested that some ELT course books can make learners feel alienated as a result of the cultural representation, which can result in disengagement from learning the language (Gray, 2000). However, this view is not in line with intercultural approaches which emphasise on comparison and cultural mediation as opposed to inclusion of facts about culture, whether this be the target culture or the students' own culture.

As previously mentioned in chapter 2, the ELI at KAU currently adopts the Cambridge English Unlimited Special Edition coursebook for all of the four ELI levels offered and, prior to the curriculum change, the ELI adopted the Oxford New Headway Plus Special Edition. The special edition coursebooks are designed to be more culturally relevant to Arab students specifically and to prepare 'Arabic-speaking learners to use English independently for global communication' (<https://www.cambridge.org>). However, the extent to which special edition coursebooks effectively promote a intercultural approach towards learning English has been questioned. Ahmadi and Shah (2014) looked at cultural representations in the New Headway Plus Special Edition found that there was a 'lack of intercultural understanding and improper representations of the Arab (Saudi) culture' which 'led to students' demotivation' (p. 18). The study further found that some parents of the students in the ELI of the Saudi university under study felt that the representation of Western values was 'alienating' their children 'from their own cultural values' resulting in 'eroding their identity' (ibid.). Researchers concluded that the special edition coursebook 'does not bring any significant change in the attitudes of Saudi EFL learners towards the target culture' (ibid.).

Two main concerns emerge from this view of how culture is featured in some EFL materials. Firstly, it does not appear that the New Headway Plus Special Edition coursebooks foster the development of cultural mediation as the difference between the standard edition and special edition consists primarily in the representation of more traditional gender roles and pictures representing women dressed more modestly. Secondly, the view expressed by some parents further identifies a lack of understanding of the purpose behind learning about cultural practices and beliefs which are different from that of the students' own culture. The concept of cultural mediation is again central here to ensure students develop a more critical view of

culture. Sulaimani and Elyas (2018) identify the notion of fear of losing one's identity as a result of learning about practices and beliefs of another culture as a concern and argue that glocalization proposes as a solution to the public skepticism on foreign cultural aspects. Robertson (1995), defined glocalization as 'the simultaneity– the co-presence of both universalizing and particularizing tendencies' (Robertson, 1995, p. 25).

Some EFL materials have moved towards a more glocal perspective on culture, for instance through the 'adaptation of international versions of EFL textbooks in order to reflect students' lifestyles, local cultures and traditions in a modern way appropriate to the twenty-first century (Holliday, 2005). However, it has been suggested that some adaptations of EFL materials, such as the New Headway Plus Special Edition do not portray a glocal perspective, because of the unbalanced representations of women in Saudi society (Sulaimani and Elyas, 2018). Khondker (2004) distinguishes between 'hybrid' editions and 'glocalized' ones and argues that glocalized materials should 'blend or adapt two processes where ideologies and cultures can act as a fusion of both local and global' (in Sulaimani and Elyas, 2018, p.73). A hybrid product, on the other hand, while addressing conditions of the market, does not specifically involve local components (ibid.). Furthermore, the importance of developing a critical perspective on ways in which cultures are presented in EFL materials or tests should also be stressed. Tomlinson (2012) argues that it is inevitable for coursebooks to communicate a view of the target culture(s) that represents the worldview of their producer. This can be problematic particularly where the coursebook is 'revered' as the 'authority' as there is a risk of its users 'uncritically accepting its views' (ibid., p. 165).

It has further been argued that foreign language teachers are faced with the challenge of preparing students to enter a global economy, in addition to teaching students to interact in a language different from their own (Kramsch and Yin, 2018). On the one hand teachers are expected to teach language as a 'cultural icon of a national unity' and on the other as a 'tool of global communication'. Kramsch and Yin (2018) argue in favour of a view of the global economy as a 'glocal contact zone', in which 'local cultural and historical interests confront global economic necessities' (p. 34).

The worldliness of English thus refers both to its local and global position (Pennycook, 1994). However, with the English being used as a global lingua franca across a wide range of diverse cultural contexts ‘a correlation between the English language and a particular culture and nation is clearly problematic’ (Baker, 2011, p.62). All English language teachers are affected by globalization; this has implications for the choice of materials to the choice of which variety of English is more appropriate (ibid.). As Kramsch (2009, p.190) points out, traditional assumptions in language teaching such as the existence of ‘pure standard national languages’ and ‘monolingual native speakers’ have all become problematic in an ‘increasingly multicultural and multilingual’ world. Kramsch’s notion of ‘third place’ (see Kramsch, 1993), offers a different way of looking at the learners’ own language (L1) and culture (C1) and the foreign language (TL) and culture studied (TC). In this view, ‘communication takes place in a sphere that is neither part of a first language/culture (L1/C1) or a target language/culture (TL/TC)’ (cited in Baker, 2011, p.64) and it is ‘freer’ and more ‘fluid’. Byram (1997) proposed the concept of an ‘intercultural speaker’ as an alternative to the native speaker model. This alternative is arguably a better model for EFL learners, as it ‘acknowledges the importance of identity and affiliation’ within the negotiated communication with ‘no one interlocutor providing the norms or ideal model to which the other has to conform’ (Baker, 2011, p. 65).

The question of cultural content has further emerged as problematic in the context of EFL testing. It is argued that a glocalized adaptation to testing in EFL could address issues of cultural relevance and provide a better balance between the global and local spheres. Examiners and test developers should be aware that test items may present some difficulty for culturally diverse test takers. Taking into consideration the culture of the test taker can, arguably, also help ensure fairness and accuracy of the test results (Marsella and Leong, 1995; Rossier, 2005). Where the cultural content of a test is adapted, however, there is always the challenge of keeping a careful balance between the original form and wording (linguistic equivalence) and any modifications (adaptations) made to the content of the test (Iliescu, 2017). If too much of the test is modified, the adapted version may arguably be considered another test. Iliescu (2017) argued that as principle ‘as little as possible’ of the content should be adapted and, where adaptation takes place, it should follow the principle of ‘intent over



content’ (p.519) so that the intention or aim of the test does not change as a result of the content adaptation.

### **3.7. Assessment and testing**

Assessment is an essential process in education and in recent years has undergone considerable development and seen a number of reconceptualization of ideas (Gipps, 2012, p.1). A key change concerns the paradigm shift from ‘psychometrics’ to ‘a broader model of educational assessment’, or rather from a ‘testing and examination culture’ towards an ‘assessment culture’ (ibid.). While the shift towards an assessment culture is visible in some learning context, for example in British primary and secondary education or through the use of formative assessments, tests and exams are still extensively employed in a wide range of learning contexts, including HE. There is, however, a much wider range of assessment in use now, compared to thirty years ago. As Kuhn (1970) suggests, a paradigm shift generally occurs when the old paradigm is unable to deal with an outstanding problem; in the case of assessment the psychometric model was viewed as no longer adequate for the assessment needs of modern-day education, which resulted in a shift in paradigms and a development of alternative forms of assessment (Gipps, 2012, p.1).

Theory on assessment has also explored its relevance to the teaching practice. The term *assessment literacy* refers to the ‘understanding of educational assessment and related skills to apply such knowledge to various measures of student achievement’ (Stiggins, 1991 in Xu and Brown, 2016, p.149) and it is ‘increasingly being recognized as an integral part of teacher professionalism’ (Abell and Siegel, 2011; Brookhart, 2002; Engelsen and Smith, 2014; Schafer, 1993; Stiggins, 1995 in Xu and Brown, 2016, p.149).

In education, the terms *assessment* and *testing* are at times employed interchangeably, however, a distinction should be made between the two. While *testing* generally includes specified tasks to be performed by students, which are drawn from a broader field of content, *assessment* includes a much larger variety of procedures, often aimed at monitoring students’ achievement and progress or the effectiveness of a programme (Chase, 1999, p. 44). A key

issue relevant to this study is the role that contextual factors play in influencing test takers' performance. Maddox (2015) argues that the testing situation is often a 'highly regulated social space' in which test takers are expected to 'perform the prescribed roles that are elaborated in assessment frameworks', yet tell us very little about 'how testing situations play out in practice' (p.431). Context-related factors, such as educational experience or computer literacy, can therefore also influence test takers' performance.

The OOPT, central to the present study, has a specific aim: to assign students to a language course adequate to their level of proficiency in English. It hence falls into the *testing* category, as it is testing specific skills (listening and reading), and therefore does not aim to give a broad overview of students' competencies across the four skills, nor is it employed for a formative purpose. Placement tests, such as the OOPT, are designed for specific purposes and therefore may differ from other tests (e.g. achievement tests) in design and ways in which they are employed by educational institutions. Green and Weir (2004, p.467) suggest that 'placement testing is an area that has received comparatively little attention in language testing research'.

### **3.8. Types of tests and their purposes:**

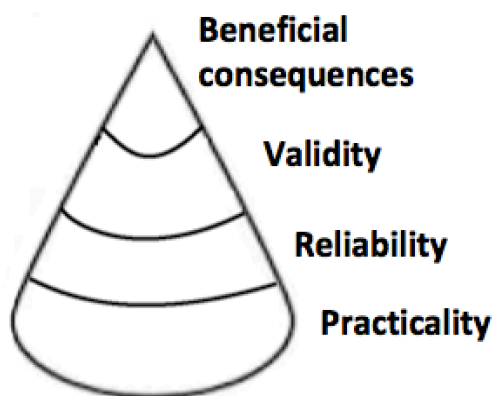
The primarily goal of testing is, arguably, to establish ways of making decisions about students' competence according to scores, in a fair way. Language testing can provide us with an accurate reflection of a student's competence and performance in a particular area. There are different purposes associated with language testing and their unique purpose can affect what test designers decide to include or exclude from a test as well as how the test will be employed in a given educational context.

With regards to the usefulness of tests, Cambridge ESOL described four essential criteria of usefulness in a given test:

- *Validity* - the extent to which test scores are a true reflection of underlying ability
- *Reliability* - the extent to which test scores are consistent, accurate, and therefore dependable for decision making
- *Impact* - the effect (preferably positive) a test has on candidates and other test users, including wider society
- *Practicality* - the extent to which a test is practicable in terms of resources needed

These four criteria play a significant role in ensuring that tests provide a useful tool for teachers to obtain valid and reliable results. Green (2014) similarly theorised four essential qualities that make ‘useful language assessment systems’ (p.58), illustrated in the figure below.

**Figure 5. Four qualities of useful assessments (Green, 2014)**



Reproduced from Green, 2014, p.58

Green (2014) claims that these are key aspects of tests that ‘users look for when choosing an assessment and that educators look for when evaluating an assessment system’ (p.58).

Equally important is the emphasis placed on test purposes. Coombe et al., (2007) points out that one of the most important tasks of any test writer is to determine the purpose of the test they intend to design. If the purpose is clearly defined, this can aid the process of selecting the right type of materials and tasks to be used in the test. The table below described five common test types along with their main purpose, as described in Coombe et al., (2007).

**Table 2. Common test types**

Common Test Types	
Test Type	Main Purpose
Placement tests	Place students at appropriate level of instruction within a program
Diagnostic tests	Identify students’ strengths and weaknesses for remediation

Progress tests or in-course tasks	Provide information about mastery or difficulty with course materials
Achievement tests	Provide information about students' attainment of course outcomes at end of course or within the program
Standardised tests	Provide measure of students' proficiency using international benchmarks

(Reproduced from Coombe et al., 2007)

Harlen (2007) revealed that there are two purposes for assessing students. The first one is 'formative assessment', that is 'the assessment is carried out in order to help learning' (p. 15). The second purpose for assessing students can be described as 'summative assessment'. This kind of assessment serves a different purpose where it 'is carried out for the purpose of reporting the achievement of individual students at a particular time' (Harlen, 2007, p. 16).

It could be argued that placement tests, however, do not fall in either category as they serve a different purpose: assessing very specific skills in order to assign students to a particular course. Placement tests, however, could be regarded as having a formative function since the accurate placement of students into a level appropriate to their ability can facilitate effective learning. In this way they differ from language proficiency tests such as the International English Testing System (IELTS) or the Test of English as a Foreign Language (TOEFL), which aim to assess a students' competency in the language across the four skills. Bachman (2007) reviewed language testing practices over the past five decades and categorised them into seven approaches:

- skills and elements
- direct testing/performance assessment
- pragmatic language testing
- communicative language testing
- interaction-ability (communicative language ability)
- task-based performance assessment
- interactional language assessment.

While Bachmann (2007) presents the seven approaches as distinct, there is some overlap among them. *Skills and elements* emphasises the assessment of the four skills (listening,

speaking, reading and writing), as well as key aspects or elements of language (phonological, grammatical, lexical and cultural knowledge). *Direct testing* or *performance assessment* places emphasis on the use of tasks and simulates real-world language use. *Pragmatic language testing*, on the other hand, does not view language ability as comprising of distinct skills and elements, but rather views language as a unified general ability. Assessments include procedures such as an interview, composition, dictation or a cloze passage. *Communicative Language Testing*, according to Bachman (ibid.) draws upon functional linguistics, sociolinguistics, psycholinguistics and language teaching. It looks at students' knowledge about the language and ways in which it can be applied appropriately in communicative contexts. *Interaction ability* draws upon *communicative language testing* but assumes that language ability is comprised of interrelated sub-competencies that play a role in language use situations. As a result of increased interest in the use of tasks, two orientations towards *Task-based performance assessment* have evolved. The first one looks at types of tasks, while the second orientation is essentially a repetition of *Performance assessment*: tasks are situated in real-world communicative language use and performance is evaluated based on considerations of real-world situations. Unlike the first six approaches defined above, the *Interactional approach* emphasises on interaction of language ability, social contexts and communication between participants. This approach is informed by a socio-cultural perspective and recognises assessment as a type of social practice affected by multiple factors.

### **3.9. Placement or proficiency tests?**

Alderson et al. (1995) argues that 'tests tend to fall into one of the following broad categories: placement, progress, achievement, proficiency, and diagnostic' (p.11). Issues may arise when a test originally designed for one purpose (e.g. as a proficiency test) is employed for a different purpose (e.g. as a placement test). Alderson et al. (1995) suggested that while a test may be valid for more than one purpose, the validity for its alternative use 'needs to be established and demonstrated' (p.170). The problem of using standardised admission tests scores for purposes other than originally intended is unfortunately under-researched (Kokhan, 2013, p. 471). There are very few studies which explore the misuse of standardised tests for

placement purposes, and these primarily look at the faculty's reflections regarding the consequences of using test scores for placement (ibid.).

Morante (1987) suggested that admission tests could not substitute for placement tests since they were designed to pursue different goals. Morante (1987) found that the greatest concern raised by the faculty, was that admission tests did not measure what they wanted their students to know. Researchers, generally agree that employing multiple measures of testing is preferable to tests which focus on specific skills (Wesche et al., 1996, and Wall et al., 1994). They further argue that where a single method is adopted, the test may be biased in favour of learners, who are better at taking tests of that kind. This issue can be addressed by introducing different test methods. A balance should also be reached between test reliability and relevance to realistic language use. While indirect tests, which may measure grammar, vocabulary or dictation skills, may be more reliable, they fail to reflect realistic communication (Green, 2012, p. 164). Direct tests, on the other hand, which may include reading comprehension, essays or interviews, are better able to reflect the kind of tasks students will need to carry out within the language classroom and beyond. However, direct tests may be less reliable and therefore the extent to which they are able to measure students' performance is restricted, thus a combination of methods should be employed (ibid.).

### **3.10. Placement tests**

Placement tests have been employed in education for quite some time, with the first one being written in the USA at Harvard between 1874 and the end of World War 1, where tests were used to place soldiers in an appropriate military position (Elliot, 2005). Later, when the US immigrant student population grew, placement tests became important to assign students into appropriate language learning programmes. The first English language programme was created at the University of Michigan in 1941. Since then, the use of English language tests has become the norm across many institutions both in English-speaking countries and worldwide (ibid.).

Many researchers have emphasised the importance of placement test in placing students in suitable courses that match their needs (Casazza and Silverman, 1996; Maxwell, 1997; McCabe, 2000). This initial stage can help support students in their future studies (ibid). Some studies have shown that in institutions where placement is not mandatory, many students do not follow placement recommendations (Latterell and Regal, 2003; Jue, 1993). This can result in students enrolling in courses that are not suitable for their level. Studies have found that students generally perform better if enrolled in the recommended course compared to students who enrol in courses beyond their level of ability (Callahan, 1993; Isonio, 1992; Marshall and Allen, 2000; Mercer, 1995; Sturtz and McCarroll, 1993). As a result, university departments want to ensure students are placed accurately.

Green and Weir (2004) suggest that there are two fundamental roles of placement tests, which emerge particularly from the literature on testing (see Harrison, 1983; Hughes, 1989; Bachman, 1990; Heaton, 1990; Alderson et al., 1995; Bachman and Palmer, 1996; Brown, 1996).

Green and Weir (2004, p.467) argue that these two fundamental roles are:

- To aid the creation of student groups of homogenous language ability
- To probe students' mastery of course content

In most cases it is the first role that predominates because of practical considerations, as well as the educational need for homogeneous ability groups (Bachman, 1990; Brown, 1996).

Wall et al. (1994) also distinguish between two different types of placement tests:

- Placement tests with a proficiency orientation (such as the OOPT)
- Placement tests developed to specifically reflect the nature of language courses

In similar vein to the two fundamental roles theorised in Green and Weir (2004), the first type 'has no direct relationship to the content of the language course into which students are placed' (Kokhan, 2012, p.292), while the second type is designed to reflect the content of the

language courses. The first type may also be considered an ‘institutional version of a proficiency test’ such as the TOEFL (ibid.). This distinction helps illustrate the different types of placement tests available and their different purposes. In the case of the OOPT, the test does not specifically reflect any particular content, rather it has a proficiency orientation. In-house placement tests, on the other hand, are often designed to reflect the course content.

While placement tests may be very helpful to ensure students take courses, which are designed to support their language needs, one of the issues that may emerge is misplacement into a course, which is either too easy or too challenging for the learner. Bachman (2004) refers to two types of misclassification errors: (1) False positive (where students are placed in a course above their actual level) and (2) False negative (where students are placed into an inappropriately low level). Both of these misplacement errors ‘can lead to negative consequences such as inefficiency of teaching and learning and dissatisfaction among both teachers and students’ (Shin and Lidster, 2017, p.358). Therefore, it is crucial that considerable effort is made to minimise misplacement errors of this kind and ensure students are accurately placed in a course that best matches their ability.

However, as Bachman and Palmer (2010) suggest, there is a misconception about testing, which lies in the belief that ‘there is a single ideal test for any given testing situation’ (Bachman and Palmer, 2010 in Crusan, 2014, p.5). Similarly, Ruecker (2011) emphasises that ‘no placement system will ever be completely effective due to the complexity of student preferences and abilities’ (p.111). He and Tymms (2005) point out that ‘two of the most important factors affecting an examinee’s performance during a test are his/her ability and characteristics such as the level of difficulty and the discriminatory capability of the test items’ (p.420-421).

Furthermore, every test has some uncertainty regardless of how well it is designed. This uncertainty is referred to as ‘error of measurement’ and it can also potentially lead to misplacement. The better the assessment the less the uncertainty but it never goes away. In education, measurement error can be defined as either ‘(1) the difference between what a test score indicates and a student’s actual knowledge and abilities or (2) errors that are introduced when collecting and calculating data-based reports, figures, and statistics related to schools and students’ (edglossary.org). Measurement error is regarded as a rather minor issue if test



results are not used to make very important decisions. However, as the stakes attached to test performance increase, measurement error becomes a more serious issue, since test results may have several consequences (ibid.). Measurement error is one reason why several test developers and testing experts advise against employing only one test result to make important educational decisions (ibid.).

### **3.10.1. Purposes of placement tests in language programmes:**

The use of placement tests in language programmes, often has the purpose of ‘matching a learner to the most suitable class or group. Often this is done on the basis of language ability as measured by a test’ (Green, 2012, p.164). These tests are most frequently employed when students enter an educational institution where teachers have very limited evidence of students’ abilities (ibid.). According to Green (2012), placement tests may also be used ‘to inform decisions about whether learners should be exempt from a course; at some universities international students will not need to attend language support courses if they score above a threshold level on a placement test’ (p.164). The use of placement tests can thus make teaching and learning more efficient by grouping students with similar language abilities (ibid.).

Green (2012) theorised two interpretations of the word ‘*place*’. The first interpretation implies placing students in their appropriate level of ability, while the second interpretation refers to placing students in relation to their knowledge of what they will be taught in the programme. The second interpretation of the word ‘*place*’ involves placing each student in relation to the institution’s learning materials; it examines how much students already know and focuses mostly on what they need to know on their future. This interpretation implies using tests that are closely associated with the content of the program; ‘in effect a variation of a terminal achievement test that is given before the course rather than at the end’ (ibid., p. 165). This type of test might be called ‘pre-achievement’ (Wall, Clapham, Alderson, 1994, p. 322).

Furthermore, Green (2012) makes reference to both language aptitude tests and general proficiency tests. He argues that while proficiency tests are widely used in context, with students of different levels of ability, language aptitude tests ‘can be used to discriminate between learners before they begin studying a language’ (p. 165).

The main objective of English placement test is to determine students’ level in language skills, that is, reading, writing, speaking and listening. According to Crusan (2014, p. 1), ‘a good placement test discriminates between test takers, allowing for their separation into groups; and this separation is useful when it comes to placing students in classes’. In addition, White (1994, p. 13-14) added that:

A good placement test is institutionally efficient, since it arranges students in convenient teaching groups, and is also valuable for teachers, since when the test is driven by the curriculum, it gives them relatively homogenous groups to work with.

Baker (1989) noted the importance of content familiarity and argued that teaching groups of students will be more effective if they share ‘the same degree of familiarity (or unfamiliarity) with material which will form the content of the course’ (p.101). This will allow students to master the material ‘roughly at the same time for each learner then they can move through the course together’ (p.101). He further noted that ‘ancillary skills and knowledge’ are needed in order to fully participate in a course. Students are thus required to have a certain level of ability in order to enable them to follow instructions, where ‘exact level of these abilities is not as important as that they should be developed to a similar degree in all the members of the group’ (ibid.). Another purpose of using placement tests, according to Crusan (2014), is to ensure that students do not repeat the same materials over and over again.

There are however, challenges concerning the administration of placement tests, for instance with regards to students’ computer literacy, familiarity with the test or over-preparation. These issues can affect test performance and thus result in a mark that does not represent an accurate picture of the student’s ability level.

### **3.10.2. What do placement tests measure?**

Reports from Community College Research Center (Hughes and Scott- Clayton, 2011; Scott- Clayton, 2012) have also raised questions about the validity of placement tests. Morante (2012) argues that the above-mentioned authors all repeat a common mistake made too frequently in the literature. They attempt to ‘assess the validity of placement tests by measuring the predictive validity, specifically correlating measures of achievement, such as placement test scores, with student success measures, such as course grades or GPA’ (p. 28). It has been previously pointed out that there is a clear misunderstanding of what a placement test is supposed to measure.

A placement test is a measure of achievement and, therefore, using predictive validity is the wrong approach to evaluate a placement test. A placement test is a basic skills achievement test that measures skills proficiency (e.g., in reading, writing and / or mathematics). The purpose of a placement test is to assist entering college students to select appropriate beginning courses aligned with their skill set. A placement test is not a test that should be used to predict success in college courses or in college itself.

(Morante, 2012, p.28)

As Morante (2012) suggests, placement tests are designed to assess skills or the proficiency of students at the time they take the test and these results should be employed as an integral part of the counselling/ advising process to place students in beginning college courses. The following section discusses the different purposes of placement tests used in foreign language programmes.

### **3.10.3. Challenges in placement testing:**

According to Crusan (2014), there are several challenges concerning placement tests. For instance, students may underestimate the consequences and impact of placement test preparation. She argues that when students take a placement test for the first time, they might be unaware of either its purpose or its expectations. She added that ‘some may fear that acceptance will be affected, while others may discount the tests altogether, failing to comprehend the high-stakes nature of placement’ (Crusan, 2014, p.2). Social practice also needs to be taken into consideration when exploring the challenges of placement testing. Cooper and Denner (1998) suggest that ‘the concept of culture has come to the forefront of

social-science and policy to address issues of human diversity in psychological processes and performance' (p.2). Debates about the role of culture have led to a shift towards research, which is 'directly applicable to social problems'; a key challenge concerns reconciling 'community-specific applications with broader theories that guide research' (ibid.). Therefore, it could be argued that theories on language testing should be interpreted by taking the socio-cultural context into account. Ercikan and Solano-Flores (2016) point out that educational assessment has become 'the focal point of educational practice' and this has had both positive and negative consequences. They argue that assessment often guides both curriculum and instruction and highlight the importance of the cultural dimension.

The same sociocultural contexts that affect student performance on tests play an important role on how assessments are perceived and used in society

(Ercikan and Solano-Flores, 2016, p. 497)

Accuracy of scores in placement tests also needs to be taken into consideration. Crusan (2014) claimed that, 'inaccurate scores can result in misplacement of students, which in turn may affect their learning experience' (p. 2). While many tests are good at discriminating between the very high and the very low levels, they are not as good at discriminating among the problematic middle levels (Crusan, 2014, p. 2).

Bernhardt et al. (2004) identify a further challenge, which may affect performance in placement tests. They argue that the 'technical qualities' of a placement test could also play a role:

A placement test even with precise technical qualities (e.g. high reliability, substantial construct validity) that does not appropriately place students into courses will affect instruction negatively by causing confusion, disorder, and frustration.

(Bernhardt et al., 2004 p. 360)

Green (2012) suggested that to lessen frustration that may occur from misplacement, students should be moved to their appropriate level. Crusan (2014) further highlights another issue with using placement tests, which concerns adopting 'scores from tests not specifically developed for placement' (p. 3). In her article, she mentioned Kokhans' (2012) study, where the researcher concluded that using TOEFL scores for placement into an ESL writing course might not be suitable to ensure students end up in the correct course.

In the USA, the majority of institutions use placement tests (NCES, 2004); according to Medhanie et al. (2012), institutions use either locally constructed placement tests or ones that are designed by an agency. There are many college placement tests on the market that are dominant in the USA. ACCUPLACER®, is one example of a placement test, which was developed by the College Board, and is adopted by more than 1,300 institutional users as an online testing system (College Board, 2009). It consists of six tests (Sentence Skills, Reading Comprehension, Arithmetic test, Elementary Algebra, College-Level Math [CLM] test, and Written Essay (College Board, 2003). Another example of a widely used placement test in the USA is COMPASS®, developed by ACT Inc. ACT COMPASS®, according to Crusan (2014, p. 3), assesses the skills of listening, reading, grammar/usage, and writing. Listening, reading and grammar are assessed using five proficiency placements: Pre-Level 1, Level 1, Level 2, Level 3, and Level 4. This is interesting as, unlike the OOPT, which only tests listening and reading, COMPASS ® tests all four skills in spite of having been designed as a placement test. It could be argued that assessing all four skills is more likely to produce accurate and valid results of students' proficiency level, however some placement tests, such as the OOPT, do not test all four skills.

Long et al. (2018) evaluated a web-based Spanish placement test used at an American university for the placement of students into Spanish foreign language courses. With regards to the skills being tested, the authors emphasise that the previously used paper-based test used to only test grammar, vocabulary, and reading, while 'the web-based test was improved to include skills that were assessed in the curriculum, such as listening comprehension' (p.141).

While some have argued that placement tests, which do not test all four skills, may fail to provide a true picture of a student's language ability, it should also be highlighted that there is an inherent difference between placement tests and other tests, such as the IELTS or TOEFL, which were designed as proficiency tests. While the latter are sometimes employed for placement purposes, they were not designed for this application.

### 3.10.4. The OOPT:

OOPT is an English language placement test, which is employed to accurately place students in a language course that meets their ability level. The time allowed for test-takers taking the OOPT varies between 50 and 90 minutes and the test-takers can decide when to submit their completed test. Tests are marked electronically, and students are given a mark out of 120. As described in chapter 2, there are marks thresholds, which correspond to six different CEFR levels. These scores can be then employed to place students into suitable language courses. The way in which marks are employed for placement purposes can vary across institutions.

With reference to Wall et al., (1994), the OOPT can be regarded as a placement test with a proficiency orientation. In the overview of the purpose of the OOPT by Purpura, testing students' language knowledge is highlighted as one of the test's key purposes:

The Oxford Online Placement Test is designed with three goals in mind: (1) to measure the language knowledge that these students have and to place them as accurately and reliably as possible into levels that align with the CEFR; (2) to provide students, teachers, and institutions (i.e., the stakeholders) with detailed information, so that they can make informed decisions about teaching and learning; (3) to provide stakeholders with an online placement test that is highly practical (i.e., inexpensive, flexible, easy-to-use, and adaptable), and reliable in its measurement of students' abilities.

(Purpura, 2007, p.2, [oxfordenglishtesting.com](http://oxfordenglishtesting.com))

The three goals outlined in the document clearly state that the test only aims to 'measure' knowledge in order to 'place' students accurately, hence it is guided by a pragmatic purpose. This is further stressed later in the document, where it is stated that the OOPT is designed to 'measure' firstly 'the test takers' knowledge of the second or foreign language (i.e., their grammatical and pragmatic knowledge)' and secondly 'their ability to use this knowledge to communicate a range of meanings while listening and reading' (ibid.). Interestingly, the word 'measure' is repeatedly used to describe the goals and aims of the OOPT. This may indicate the author's deliberate effort to distance himself from terminology such as 'assess' or 'test', which may mislead educators in understanding the purpose for which this test was designed. While not stated explicitly, the use of the word 'measure' might refer to the psychometric use by the OOPT called 'Rasch measurement'. Rasch measurement allows the difficulty of a test to be separated from the ability of the test-takers as well as creating an equal interval scale (Tymms, 2013, p.2). The Rasch model measures the 'item difficulty and

person ability on the same linear continuum' in order to provide an 'objective reality' (Wright and Stone, 1979 cited in He and Tymms (2005, p. 421).

While the OOPT is designed as a placement test to be employed internationally, the test incorporates a significant number of idiomatic expressions used in British English. This aspect of the test could be viewed as problematic, particularly for students who have not had the opportunity to spend a considerable amount of time in the UK or are otherwise unfamiliar with idiomatic language use. A study carried out at the University of Calgary, in Qatar (UCQ), found that while the OOPT provided several listening dialogues, which were realistic, it also covered a wide range of situations that local students were unlikely to encounter (Brooke et al., 2012). The investigators felt that the test would be more suitable for students who wished to study in an English-speaking country. While the OOPT's website claims that the test is designed to evaluate how learners use their knowledge of grammar and vocabulary to understand authentic communication, Brooke et al. (2012) found that 'the test has a very heavy emphasis on idioms, such as 'throw my hat in the ring' and 'call it a day.' Some idioms even stumped the faculty researchers, for example, 'turn up for the books.' The researchers pointed out that perhaps these idioms were 'specific to British English' (p. 16). Student researchers at UCQ also found other British expressions, such as 'straightaway,' 'got sacked' and 'shared a flat' confusing, as they were less familiar with British English. The cultural context related through the use of expressions further presented a challenge. For instance, the concept of 'sharing a flat' was difficult for the student researchers to grasp, as they struggled to interpret the relationship between the speakers, being unfamiliar with this living arrangement. Furthermore, British accents in the listening sections increased the difficulty of comprehension, resulting in the researchers feeling rather demoralised by the test. Brooke et al. (2012) concluded that since English is primarily used as an additional language at UCQ, the majority of the student population will have had minimal exposure to British idiomatic expressions and will probably have less need for them in the classrooms. With reference to this study, 'the OOPT was found to be less favoured than other placement tests' (Brooke et al., 2012 cited in Sinlapachai et al., 2016, p.67). The placement test included content, which students were unlikely to have encountered (Sinlapachai et al., 2016).

Sinlapachai et al. (2016) makes reference to a study carried out by Berthold (n.d.), which explored the use of online standardised tests as indicators of English proficiency, found inconsistencies when comparing the paper-based Oxford placement tests and the OOPT. The study compared scores in the 'Use of English' section and found that the two tests 'did not provide consistent CEFR levels to students' (Berthold, n.d. cited in Sinlapachai et al., 2016, p.67). Another inconsistency was found when comparing Cambridge English Placement Test (CEPT) and OOPT scores. Sinlapachai et al. (2016) found that more than 50% of students received higher CEFR levels from the CEPT compared to the OOPT across all levels. However, the study identified a positive relationship between OOPT and CEPT scores significant at  $p < .01$ .

Berthold's study further identifies other issues relating to placement tests. Firstly, he argues that the main component of the placement test, whether paper-based or computer based, is problematic because it is composed of multiple-choice questions. This inevitably means that guessing is involved. Sinlapachai et al. (2016) also looked at test-takers' performance in the OOPT. The aim of their study was to investigate whether the CEPT and the OOPT could provide an accurate indication of CEFR levels. In their study, they found that students were unwilling to take the test because they had not been adequately informed of the purpose. Some of the students who took the test also commented that the content was quite different from what they had learnt at school, in particular with regards to the 'vocabulary and idioms' (ibid. p.69). Furthermore, the study's findings indicated that the CEPT and OOPT did not provide consistent CEFR levels, although there was a positive relationship between the scores of both tests. Sinlapachai et al. (2016) point out an important criticism of both the CEPT and the OOPT. They argue that there is a mismatch between the skills measured in these placement tests and the much broader range of skills included in the CEFR. While the CEPT and the OOPT only test receptive skills, the CEFR also makes reference to students' productive and interactive skills. Yet, both placement tests equate a certain score with its corresponding CEFR level. They argue that the CEPT and the OOPT were designed for placement purposes and were therefore not intended to measure students' proficiency.

While these studies explored aspects of the OOPT and other placement tests, such as the CEPT, in their own institutions, it is evident that more research is needed to explore specific



aspects of this test and its suitability for students who are generally unfamiliar with British culture and idiomatic language use. Previous research thus also suggests that students might perform differently on a paper-based version of the test compared to the online version. Computer-adaptive tests take less time to indicate the level achieved by a test-taker because, as Brown (1997) suggests, the test-items are individualised for each test-taker. Computer-adaptive tests have the potential to provide more accurate and reliable information in a shorter amount of time (Merrell and Tymms, 2007, p.27). CATs are different from a traditional test in that they are ‘challenging but within reach and shorter’ (ibid.). Future research could explore whether these findings might be related to computer literacy or other aspects of the test.

A further study carried out by a faculty member at KAU explored students’ achievement in an English language placement test within the university, although the name of the placement test is not mentioned. Mahmoud (2014) carried out a needs-analysis exploring the views of students and faculty members (across different faculties within the university) with regards to the foundation year programme. While the study did not focus on the placement test, the article makes mention of the percentage of students who were placed in the four different ELI levels. He found that out of 4360 students who took the placement test 71.1% were placed in level 1, followed by 16% in level 2, 9.2% in level 3 and only 3.67% in level 4. As can be inferred from the data, the overwhelming majority of students were placed in the lowest level.

While Mahmoud (2014) did not explore possible causes of low results in the placement tests, he suggests that this could be due to the ‘huge gap between the school education and the university education’ in Saudi Arabia (Mahmoud, 2014, p.337). This finding is particularly relevant to this investigation as it helps illustrate some of the challenges that ELI teachers may face as a result of students’ weak language skills upon admission.

### **3.10.5. Relating test scores to the CEFR:**

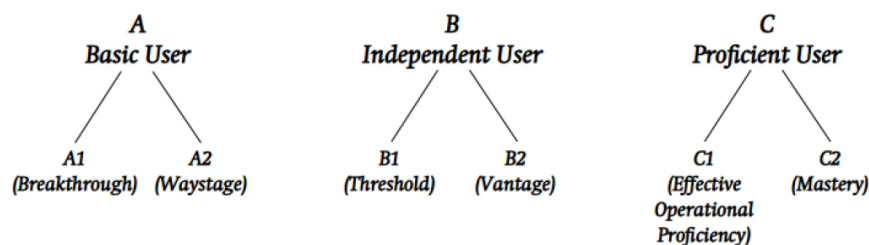
The CEFR was developed between 1993 and 2000 (Goullier, 2006), as a document outlining the progression of language proficiency and abilities across a range of learning contexts:

[The CEFR] describes in a comprehensive way what language learners have to learn to do in order to use a language for communication and what knowledge and skills they have to develop so as to be able to act effectively.

(CoE, 2001, p. 1)

The CEFR subdivides language proficiency and ability into six Common Reference Levels ranging from A1 Breakthrough to C2 Mastery (see Figure 6).

**Figure 6. CEFR Common Reference Levels**



Reproduced from CoE, 2001, p. 23.

Each of these six levels is accompanied by a descriptor detailing the abilities of learners as they progress through the levels. The CEFR is employed to aid those involved in the learning, teaching, assessment and policy of language and reflects over six decades of work carried out by the Council of Europe (CoE) (2001), an organisation responsible for endorsing plurilingualism, linguistic diversity, mutual understanding, democratic citizenship and social cohesion across its member states (CoE, 2015; Language Policy Division [LPD], 2006). The framework has helped outline the wide range of competences developed by language learners and has been ‘extremely influential’ in the development of curricula, syllabus design and language examinations (Hulstijn, 2007, p. 663). The fact that it is not language-specific also means that the framework can be employed in a vast range of learning contexts. It should also be noted that the CEFR’s model of communicative competence no longer equates the goal of second language learning to native-like proficiency. Therefore, ‘successful communication’ is measured on the basis of varying degrees of proficiency in multiple languages (LPD, 2006, p. 5) as opposed to native-like ability. Since its publication, the CEFR has also been increasingly employed in language certification (CoE, 2001).

In 2009, the Council of Europe published a manual specifically designed to provide guidance in the process of relating language examinations to the CEFR. The primary aim of the Manual was to assist in the development, application and reporting of practical procedures in order to situate examinations in relation to the CEFR. The Manual is not the sole guide to linking a test to the CEFR and therefore institutions are not obliged to undertake such linking (CoE, 2009). However, it suggests that institutions wishing to ‘make claims about the relationship of their examinations to the levels of the CEFR’ (CoE, 2009, p. 1) may benefit from the procedures described in the Manual. The Manual offers guidance to users to:

- Describe the examination coverage, administration and analysis procedures;
- Relate results reported from the examination to the CEFR Common Reference Levels;
- Provide supporting evidence that reports the procedures followed to do so.

(CoE, 2009, p.1)

The rise of language testing seeking to link test results to CEFR levels has, however, prompted both praise and criticism. Alderson (2007, p. 661) argues that one of the problems with linking test results to the CEFR is that there is ‘no theory of comprehension that could be used to identify the mental operations that a reader or a listener has to engage in at the different levels of the CEFR’. Other criticisms highlight the ‘ambiguous’ and ‘inconsistent’ nature of the terminology employed in the CEFR (Alderson, 2007; Alderson et al. 2006). The CEFR, however, has also received a number of positive reactions. Little (2007, p. 648) argues the following:

The scale seems to offer a ready means not only of indicating the degree of communicative language proficiency confirmed by a particular test or exam, but also of comparing tests and exams with one another, from language to language as well as from country to country.

Figueras (2007) similarly suggests that the popularity of the CEFR has resulted in quality testing receiving more importance. Papageorgiou (2007) found that the undertaking of a CEFR linking project contributed to the quality of the Trinity examinations.

Linking test results to the CEFR is helpful as it gives greater meaning to test scores; test-takers and institutions are therefore also provided with an indication of the test-taker’s language ability. Davidson (2012) argues that test scores need to have meaning and illustrates the following example: ‘For instance, if a student takes a test and receives a result

of ten, then a natural question arises: what does the 10 represent?’ (p.197). Davidson (2012) argues that there are two classic answers to this question: norm referencing and criterion referencing. According to the former, a score of 10 on a language test would indicate the rank of the test taker within in a group of peers. This is determined by how many test-takers scored above 10 and how many below. According to the latter, if a score of 10 is a passing mark on a language test, this could indicate that the test-taker has mastered a sufficient number of skills within a very well-defined domain of language ability (ibid, p.197). One of the criticism of the CEFR relates to the impact it has had on teaching and learning. The CEFR was initially designed to facilitate the recognition of language credentials across national boundaries and has rapidly become institutionalised across Europe (Young, 2012). Fulcher (2004, p. 260) suggested that the impact of the CEFR and the adequacy of its underlying construct are in need of debate since,

for teachers, the main danger is that they are beginning to believe that the scales in the CEFR represent an acquisitional hierarchy, rather than a common perception. They begin to believe the language of the descriptors actually relates to the sequence of how and what learners learn.  
(Fulcher, 2004, p. 260)

### **3.11. Standardised tests and their relevance to the OOPT:**

The structure of the OOPT resembles to an extent standardised tests, as it employs multiple choice questions and provides a standard English placement to be used world-wide. The first kinds of standardised tests were mass-produced and had specific sets of instructions and evaluation methods.

Standardization testing continued to expand in the USA, particularly in the 1990’s in the field of psychology. This further led to the development of standardised achievement tests, such as those developed by Thorndike. There were mixed views on standardised tests, while some regarded this new approach towards testing as a great achievement in the development of efficient assessment practices, others such as John Dewey, feared that it reduced education to averages and percentages. These kinds of tests were also employed by the US Army during World War One. This was perhaps the most influential historical event, which resulted in a widespread use of standardised testing in the United States (United States Congress. Office of Technology Assessment, 1992).

While standardised tests have become increasingly popular, questions regarding the educational objective of this type of testing have arisen. Gipps (2012) points out that one of the problems with these tests is that they aim to measure knowledge rather than support learning. She further argues that ‘the huge quantities of multiple-choice standardised tests in the USA, and the formal written exams in the UK – have had [...] unwanted and negative effects on teaching and the curriculum’ (p.3). Hence, while standardised multiple-choice tests may offer learners a fair assessment and lessen the burden of marking on teachers, they also have their drawbacks. As with standardised tests, the extent to which a multiple-choice placement test, such as the OOPT, can support learning, is questionable. Al-Sadaawi (2010) and Al-Sadan (2000) have described assessment tests as being gap filling, multiple choice, true or false and essays, and Taras (2005) suggests that these do not extend students’ abilities to use their learning out of context. However, with regards to placement tests, students generally do not see their test after it has been marked, nor do they receive any feedback. Placement tests are, therefore, inherently different from achievement or proficiency tests with regards to the purpose for which the test is designed and the extent to which they inform learning.

### **3.12. Computer Adaptive Tests (CATs):**

One of the characteristics of the OOPT is that it is a computer adaptive test (CAT) and thus rather than having a set number of questions, which are the same for all test-takers, the computer selects subsequent questions on the basis of performance on previous ones.

Since 2001, large-scale assessment and computer technology have evolved rapidly (Gierl, Lai, Li, 2013, p. 188). Nowadays, computers play an essential role in language testing. Most institutions around the world have incorporated computer technology in their teaching, learning and testing process. He and Tymms (2005) have mentioned different kinds of computer software systems that have been developed to conduct tests and analyse its results. Such as, systems delivered over the Web (e.g. CASTLE, Hot Potatoes, WebTest, TAL, TOAI, Questionmark TM Perception TM, i-assess) (p.420).

CATs are based on item response theory that was formally introduced by Lord in 1980 (Wainer, 1990). In the CAT, the computer estimates the candidate’s ability level and after

rough estimation, it selects the most appropriate item that matches his or her ability level. According to Alderson (2000), ‘if the user gets an item right, they can be presented with a more difficult item and if the user gets the item wrong, they can be given an easier item’ (p. 596). In other words, in CAT the computer adjusts the appropriate item to be delivered according to the student’s success or failure on the previous item. In this case, if a student ‘fails a difficult item, s/he is presented with an easier item, and if s/he gets an item correct, s/he is presented with a more difficult item (Alderson and Banerjee, 2001, p. 225). This means that ‘no two test takers are likely to face the same set of items, assuming that the item pool is large enough’ (Fulcher, 2010, p. 205). He and Tymms (2005, p. 420) described the process of CAT as follows:

in computer adaptive test, for a particular examinee, the items, drawn from an item bank containing items that have been calibrated using an IRT model, are targeted at his/her ability level

Luecht, Champlain and Nungester (1998, p. 29) have noted that, by targeting the appropriate item for each student:

We can maximize the obtained measurement information for each test taker and improve the reliability of score estimates or classification decisions. Maximising measurement information minimizes measurement errors. CAT is therefore said to be *statistically efficient* either because: (i) the test length can be reduced on average for many examinees with no loss in measurement precision or decision accuracy, or (ii) the precision of scores or decision accuracy can be increased for a fixed or average test length.

CAT technology is unlike the traditional paper-pencil test, where all students receive the exact same questions. In CAT, each examiner receives different items according to his or her ability level. There are some advantages and disadvantages of using the CAT applications for testing purposes. Merrell and Tymms (2007, p. 35) suggested one advantage of CATs is that:

This system has advantages over a traditional pencil-and-paper group assessment in terms of the ease and time for administration, the ability to provide items appropriate to the ability of individual pupils, which gives more reliable results for pupils at the extremes of the normal population, and a reduction in administration, marking and analysis.

Both Alderson and Banerjee pointed out that using CAT has its advantages. Firstly, students are not presented with items that are too easy or too difficult, but they are presented with items that are adapted to their ability level. Secondly, CAT ‘are typically quicker to deliver, and security is less of a problem since different candidates are presented with different items’

(Alderson and Banerjee, 2001, p. 225). Furthermore, the shortened testing time in CAT can help to avoid fatigue or other unintended sources of variance that may reduce the validity of score interpretations (Katz and Gorin, 2016).

Using adaptive test format for testing is shorter and less time-consuming than a conventional paper-based exam, because each candidate answers questions that are relevant to his or her language ability. In addition to the previously mentioned advantages of using computer-based tests (CBT) and computer-assisted assessment (CAA) over traditional paper-and-pencil-based testing, He and Tymms (2005, p. 419) added that: ‘these include immediate, unbiased and accurate scoring and feedback, increased efficiency, convenient individualised administration and improved test security’. In similar vein, an advantage of adaptive tests over conventional tests is that:

Adaptive tests are designed to improve measurement over conventional tests by improving the efficiency of test administration—by administering the minimum number of items necessary to measure each examinee—and by controlling the precision of measurement.

(Weiss and Shleisman, 1999, p. 129)

On the other hand, using CAT does have some disadvantages. According to Boyd (2003) and Eggen (2001), security is one disadvantage of using CAT tests as some students could memorise test items that they have been exposed to and share them with other students who plan to take the test. To address this issue, Huang, Lin and Cheng (2009) suggested building a large item bank to control test item exposure.

Thus, in spite of the advantages of CATs, these are complex tests to design:

They are not very widely used in institutional placement testing because they are very complex to compile (it is difficult to ensure that all test takers are provided with a similar range of content), require quite sophisticated computing and statistical knowledge to construct, and need large numbers of questions to cater for the possible routes through the test.

(Green, 2012, p. 169)

The measurement model frequently used in computer adaptive testing is Rasch’s model (Bond and Fox, 2007). In this model:

All test takers are placed on an ability scale, with 0 as the mean. Items are also placed on the same scale. However, while the level of the test taker is interpreted as ability, the level of the item is interpreted as its difficulty. The lower the number, the easier the item is. Similarly, for the test taker, the lower the number, the lower the ability estimate is.

(Fulcher, 2007, p. 206)

We can conclude that while CATs may offer advantages for learners, there are also drawbacks of adopting this particular kind of test. While allowing learners to answer questions, which are tailored to their level of ability, may be a benefit, it could be argued that it may not give all learners an equal opportunity since the items may vary.

### **3.13. TOEFL and IELTS tests:**

TOEFL and IELTS are considered pre-arrival exams, however these test results are at times used for the purpose of placement into ESL courses instead of placement tests at many universities around the world (Kokhan, 2013). At KAU, if students provide evidence of having taken the TOEFL or IELTS, and have achieved a satisfactory score, they can request an exemption from taking the compulsory English language modules. In order to be exempt, students need to have achieved the equivalent of completing the level 4 module. KAU therefore does not employ TOEFL and IELTS scores for placement purposes, rather students can request an exemption on the basis of these test results.

According Morante (1987), pre-arrival tests are not a good substitute for placement tests, as each test has been designed with a specific goal in mind. Fox (2009), in a study on using standardised proficiency test scores (e.g. TOEFL and IELTS) for placement in EAP courses in a Canadian university, found that there were a number of negative consequences resulting from using standardised proficiency test scores for placement purposes.

The results presented by Fox (2009) may not necessarily mean that using proficiency test scores for placement purposes in other contexts is a bad idea. It is possible that the test scores were not working for EAP placement because the curriculum for the EAP program required a different set of abilities, and getting a certain score on the TOEFL did not necessarily mean that a student with such a score could perform well in a particular EAP class.

(Kokhan, 2103, P. 471 - 472)

The ‘International English Language Testing System’ or IELTS was designed to ‘assess the language ability of candidates who need to study or work where English is used as the language of communication’ (examenglish.com). It was jointly established in 1989 by three different international organizations: the British Council, IDP: IELTS Australia and the University of Cambridge ESOL Examination.

On the other hand, the TOEFL is a trademark of the Educational Testing Service (ETS), a private non-profit organization that designs and administrates the test. The TOEFL was



launched in 1964, and since then it had been revised several times. The TOEFL IBT, launched in 2005, is the most recent revision. Many innovative design features have been added, including ‘the use of integrated tasks that engage multiple language skills to simulate language use in an academic setting, and the use of test materials that reflect the reading and listening demands of a real-world academic environment’ (Lawrence, 2011, p.1). Both the IELTS and TOEFL have advantages and disadvantages in the way they are presented.

With regards to the use of proficiency test scores for placement purposes, Crusan (2014) argues that one of the issues in placement testing is ‘the use of scores from tests not specifically developed for placement’. She suggests that, particularly in an era of reduced funding, many ‘language programme administrators are investigating the use of standardised language proficiency test scores to place students’ (Crusan, 2014, p.3).

The ELI at KAU has favoured the OOPT over the IELTS and TOEFL because it is less time consuming and it is mainly used to place students at the appropriate level in the English modules. IELTS and TOEFL results are only used to grant students an exemption, as previously mentioned. Hence the university differentiates between the use of placement tests and pre-arrival tests and employs these for the purposes for which they were originally designed by their test developers.

### **3.13.1. Institutional use of pre-arrival and placement tests for placement purposes:**

Universities across the world employ both pre-arrival tests, such as TOEFL and IELTS, as well as placement tests, such as the OOPT to fit their purpose. Particularly in institutions where English is not the language of instruction, but students are required to achieve a certain level of proficiency, English language programmes often need to ensure that incoming students are placed into the appropriate English course. The purpose of requiring students to sit a placement test, in this case, is to facilitate teaching and to ensure that students benefit from the course. As a result, universities may choose to employ a test, which was originally designed for a different purpose, as a placement test, or may use the test results differently. As previously mentioned in chapter 2, the ELI at KAU provides a good example of how an institution decided to adapt the way OOPT results are used to place students into different CEFR levels. The ELI offers four proficiency levels instead of six and groups all students

scoring higher than 50 marks into the intermediate level (the highest offered). While this use of the OOPT results may be questionable, one needs to take into consideration the purpose for which the OOPT is employed at this particular university. At KAU, the majority of the students enter university with an elementary / intermediate competence in English, and it would therefore be unproductive to offer classes at the advanced level, which may only be suitable for a small number of students who may have had an extensive opportunity to learn the language prior to university.

Similarly, universities in the USA are also increasingly employing pre-arrival results, for placement purposes; As Kokhan and Lin (2014) point out ‘the vast majority of U.S. universities nowadays accept TOEFL iBT scores for admission and placement into ESL classes’ (p. 1). While there are apparent advantages for students and the university (such as reducing time and costs associated with students sitting a university-based placement test), research shows that placing students into ESL classes based on scores obtained from pre-arrival tests may result in misplacement. Mullen (2008) points out that the use of proficiency tests for ESL placement is being increasingly adopted across many educational institutions, however ‘proficiency tests do not test the incremental developmental changes that define learning within an EAP program’ (in Kokhan, 2012, p.294). Kokhan (2012) found that TOEFL scores ‘may not be fully suitable for placement into the ESL writing courses’ as neither the ranges of TOEFL total and subsection scores, nor a specific cut-off score (for exemption purposes) could be established (p. 306). The study also found that ‘time lag’ played a role and that a year was long enough for ‘TOEFL scores to lose their relevance’ (ibid.). Kokhan (2013) found that employing pre-arrival tests such as SAT Reading scores and TOEFL iBT, not all students were accurately placed in the appropriate ESL level. In fact, results showed that there was about a 20% chance that students with low TOEFL iBT total and TOEFL iBT Writing scores would be placed into the highest ESL level. She argues that additional on-site testing may still be required to prevent misplacement and thus the use of pre-arrival test scores may be questionable. These findings are further confirmed in Kokhan and Lin (2014), who found that ‘TOEFL iBT scores may not be a reliable predictor of ESL placement’ (p.17).

It is hence important to acknowledge that while educational institutions should strive to employ tests for their purpose and ensure that students are placed in a course, which matches their ability, as mentioned in section 3.5, no test is a perfect instrument.

### **3.14. Teaching mixed-ability groups:**

While addressing the issue of misplacement may be important to ensure students benefit from the learning experience in English language courses, what is overlooked when employing test scores for immediate placement is that there are external factors, which can influence students' performance on placement tests. These factors include students' educational background, familiarity with computers and previous experience learning English. It should also be pointed out that it is not inconceivable to teach mixed-ability courses, however, in order to differentiate effectively, teachers are required to be adequately trained (Reid et al., 1982). Reid et al. (1982) argue that successful mixed-ability teaching relies heavily on teacher skills; furthermore, in order to teach a mixed-ability group effectively, teachers often have to spend a significant amount of time developing differentiated materials (Hallam and Ireson, 2005).

While there are advantages associated with teaching mixed-ability classes, there are also numerous challenges. A mixed-ability setting can allow students to be more creative as they learn from each other. However, teaching such courses can lead to frustration, as it is very difficult to address every student's needs. Ireson and Hallam (2001) argue that instructors need to recognise that students in mixed-ability classes will have different strengths and weaknesses and will improve at different rates even with a perfect placement. Effective mixed-ability teaching, however, requires teachers to implement differentiation strategies within the classroom.

Differentiation, as an instructive approach, has the aim to maximise the learning opportunities for all students (Bearne, 1996).

Laar (1995) noted that over the years a number of approaches have been used to differentiate between students, including the following:

1. Ordering the learning complexity of the tasks set to students
2. Devoting additional time to students with special learning needs.
3. Grouping students by ability
4. Providing different levels of support material.
5. Accepting different levels of achievement and modes of presentation

(in Kerry and Kerry, 1997, p. 440-441)

Differentiation has been described as an optimum pedagogical approach (i.e. Birnie 2015; Koutselini 2006, 2008; Tomlinson 2015) since ‘it transforms instruction into a dynamic, pleasant, satisfying, self-regulated, and effective process based on students’ needs and characteristics’ Valiandes and Neophytou (2018, p. 123). However, most instructors lack the professional development, time for lesson planning, or the capacity to consistently implement differentiation (Loiacono and Allen, 2008). Instructors in mixed-ability classrooms need to be provided with continuous training in order to better manage the challenges of their diverse learners (Butterworth, 2010). Valiandes and Neophytou (2018) found that teachers may recognise that there is a need to differentiate among their students, however some believe that it would be challenging and time consuming. Some also admitted that they did not really know ‘how to translate the theory of differentiation into practice’ (Valiandes and Neophytou, 2018, p. 123).

While mixed-ability teaching remains an area to be further explored, ensuring the reliability and validity of the tests currently employed to enrol students into a suitable language course remains of key importance, particularly for educational institutions where teachers are not adequately trained to teach classes with a wide range of proficiency levels.

The following sections explore the concepts of *reliability* and *validity* in relation to testing.

### **3.15. Reliability:**

According to Anastasi (1988), reliability refers to ‘the consistency of scores obtained by the same persons when re-examined with the same test on different occasions, or with different sets of equivalent items, or under variable examining conditions’ (p. 109). Many assessment experts (Weir, 2005; Hughes, 2003; Alderson, Clapham and Wall, 1995; Bachman, 1990), advised to use three different ways for measuring test reliability:

1. test-retest reliability: giving the same test twice to the same group of students, and the scores obtained by the same persons on the two tests administrations are correlated to give a reliability coefficient
2. equivalent (or parallel) forms reliability: giving two tests that are the same but in different forms; equivalent but not identical.
3. internal consistency reliability: ‘focuses on the consistency with each other of a test’s internal elements

(Weir, 2005, p. 25-29)

Alderson (1991) argues that although test-retest reliability is considered to be the easiest measure of reliability to conceptualise, there are still some issues with the concept. To clarify, if a person takes the same exam for the second time, and the test is reliable, the score should remain the same, or almost the same. However, scores are often different due to improvement in the candidate’s competency or an increased familiarity with the test. Bachman (1990) added that a ‘fundamental concern in the development and use of language tests is to identify potential sources of error in a given measure of communicative language ability and to minimise the effect of these factors on that measure’ (p. 160). Bachman (1990) clarified that errors of measurement or unreliability can be affected by numerous factors, as previously mentioned in section 3.5. Among the factors affecting reliability are might affect the reliability lack of interest or motivation, test-wiseness, poor health and fatigue, luck or immediate pre-experience.

A further factor that may affect the reliability of a test is related to the test itself. That is, how the test is presented (if it is presented differently on different occasions), what kind of language is used and what Gipps and Murphy (1994) refer to as ‘differential performance factors’. Bachman (1990) argues that the more these previous factors are minimised, the more reliable the test could be. ‘In other words, the *less* these factors affect test scores, the *greater* the relative effect of the language abilities we want to measure, and hence, the reliability of language test scores’ (p. 160). In similar vein, it has been suggested that:

The starting point for estimating the reliability of a test is to hypothesize that each student has a ‘true score’ on a particular test...An individual’s true score on a test is simply the average score that the individual would get over repeated takings of the same or a very similar test.

(Wiliam, 2001, p. 17)

Over time, reliability was more researched than validity, according to (Wood, 1991; Black, 1998), the reason for this was that reliability was easier to account for and validity was

neglected by the examination board (Wood, 1991, p. 151). We have to bear in mind that validity and reliability are interconnected criteria. Black (1998, p.48-49, cited in Lambert, Lines, 2000, p. 12) explains this connectedness thoroughly:

A test giving scores which are reproducible over a variety of conditions (i.e a reliable test) would nevertheless be of no value if the attainment that it measures is irrelevant to (say) the learning being explored, or if that attainment is too narrow an aspect of the whole to be of any significance (i.e. an invalid test). A set of test items which clearly reflect the aims of course and can be seen to evoke the type of performance that the course aims to promote (i.e. a valid test) is of little value if it turns out the examiners cannot agree on how to mark the outcome, or if a different set of apparently comparable items would produce very different scores with the same pupils on a different occasion (i.e. an unreliable test). If it turned out that there was no correlation between a test score and the future attainment that it was designed to predict, then it could be that either or both of the reliability and the validity of the test were defective.

Wood (1991) further emphasises that ultimately validity is more important than reliability but any attempts at validation depend crucially on the reliability of the observations. If the observations cannot be trusted, then one might reach a misleading judgement concerning validity. He suggests this might be the reason why validity has been explicated much less than reliability (p. 147).

Reliability can be seen as a 'prerequisite for validity' since we cannot validate a test if students' results keep changing radically from one occasion to another (Wiliam, 2001, p.18). In addition, Wiliam (2001) added that validity is more important than reliability 'since there is no point in measuring something reliably unless one knows what one is measuring' (p. 18). Wells and Wollack (2003, p.3) further remarked that 'if the test is unreliable, one needn't spend the time investigating whether it is valid - it will not be'. The time allocated to students to complete the tasks and the number of tasks students are asked to complete can further affect the reliability of the results. Black and Wiliam (2006, p. 120) argue that 'if the time allowed for the test is very short, the sample of items will be very small. The smaller the sample, the less confidence one can have that the result for any one candidate would be the same as that which would be given on another sample composed in the same way. Thus, any examination can become more reliable if it can be given a longer time' (Black and Wiliam, 2006, p. 120). Hence one of the challenges is to be able to provide enough tasks to accurately assess students across a range of skills.

However, while both test developers and educational institutions are concerned with obtaining reliable results, as Bloxham (2012) suggests, the assumption that a 'correct' grade

exists ‘out there’, can be attributed to a techno-rational model, while a socio-constructivist model holds the assumption that grades are constructed by the individuals who are using them and cannot be looked at separately from the people and cultures that are constructing them. That does not mean that there is no such thing as a ‘true’ score, but rather that there are different points of view on this. Kappler (2004) described the techno-rational perspective on assessment as ‘a standpoint that promotes the use of quantitative data and measurement to ensure accountability’ (in Bloxham, 2012, p. 186), although this is not always the case, for example with placement tests. Other researchers have ‘challenged the notion that it is possible to make explicit the tacit knowledge involved in assessment decisions (O’Donovan, Price, and Rust 2008; Orr 2007; Sadler 2009; Shay 2005 cited in Bloxham et al., 2011). Students, for instance, may respond to essay questions differently but in ‘equally effective ways’. ‘This requires tutors to use their judgement, based on their tacit knowledge, in order to allocate grades’ (Bloxham et al., 2011, p. 657).

With regards to this study, the socio-constructivist model can help shed light on teachers’ views on assessments. This is perhaps more relevant to the continuous assessments taking place in the modular courses than the OOPT, as this is where teachers are personally involved with assessments.

Designing tests can prove to be significantly challenging. There are concerns with regards to test content and test techniques, as well as concerns about the reliability of the instrument. It is important to ensure that the content of a test is understandable and that the test techniques are effective. In the case of multiple-choice tests, for instance, Hughes (2003) identifies that ‘good multiple-choice items are notoriously difficult to write’ (p.3). Where not enough time and effort is taken in designing these tests, ‘the result is a set of poor items that cannot possibly provide accurate measurements’ (ibid.). While tests cannot account for the human element (e.g. the mood of the test taker, health or how much sleep they had the night before taking the test), ‘having clear instructions and unambiguous questions’ can ensure that the test themselves do not ‘increase variation’ (ibid., p. 4). Reliability also depends upon the way a test is scored. Composition scoring for example may prove to be unreliable if different scores are given by the same marker on different occasions or by different markers

### **3.16. Test preparation:**

It can be argued that test preparation, test familiarity, students' background and motivation all play an important role in the learning experience. Preparation for a computer-based test, for instance, requires the learner to become familiar with the interface and ways to navigate through the test. This kind of preparation should aim to familiarise learners with the types of items on the test, the various types of instructions and allow them to practise working within time-constraints (Fulcher, 2010, p.288). This can help ensure that learners do not spend extra time and effort understanding what they should be doing and how to submit their responses. Effective test preparation may reduce the chances of scores being affected by students' unfamiliarity with the way the test is designed, which, in turn, may increase the validity of the scores by eliminating the chance of construct-irrelevant variance (ibid.). However, test preparation can also have detrimental effects on learning. In particular, national tests have often led to teaching practices that focus primarily on developing test-taking skills rather than comprehensive knowledge in the subject. Teachers and students alike are sometimes involved in unhealthy, and sometimes unethical, test preparation practices (Crocker, 2005, 2006; Lai and Waltman, 2008). Gebril (2006) suggests that such practices can lead to numerous negative effects both on teaching and learning. 'When teachers and students are involved in *teaching to the test* practices, this behaviour harms the learning process' (Gebril, 2006, p. 431). Therefore, while test preparation can improve test performance, it should not guide teaching practices. Rather, what is essential, is test familiarity so that students do not underperform as a result of failing to understand the instructions or expectations of the test.

### **3.17. Test familiarity:**

Familiarity with the content and administration of the test is helpful as it can impact on students' test performance. With reference to the present study, the investigation is concerned both with teachers' and students' familiarity with the OOPT. CBTs in particular can present a significant challenge for some test-takers and therefore it is important that they are familiar with these types of tests. Bunderson et al. (1989) refer to studies of test equivalence, which identified some inconsistencies between scores obtained from CBTs compared to those obtained from paper-based tests. They concluded that lack of familiarity



with computers could affect students' performance. A number of assessment specialists have called for more studies, which explore the extent to which there might be group differences in performance on computer-mediated assessments (Chapelle and Douglas 2006; Leeson 2006; Douglas and Hegelheimer 2007). Douglas and Hegelheimer (2007) suggested that computer familiarity might be a factor for some groups of test takers. As a result, test developers should collect evidence on the characteristics of test takers (such as gender, first language, geographical region, or socioeconomic status) in order to explore whether there are significant differences in computer familiarity among the different groups.

Issues concerning familiarity in language testing are not new, in fact it has generally been accepted that test-takers should be familiar with the format and delivery of the test before taking it. Russell and Haney (1997) suggested that testing writing through a paper-and-pencil test might be increasingly unfair in situations where learners are accustomed to using word processors. In such cases a computer-delivered version of a writing test could lead to higher scores simply as a result of familiarity.

### **3.18. Fairness and test bias:**

A further challenge of test design and administration is to ensure that the instrument is fair and that it does not disadvantage any particular group of test-takers. This is of particular relevance to the present study as the investigation seeks to explore teachers' views on the OOPT, which include reference to fairness and test bias.

According to Mitchell and Cawthon (2014), the terms 'fairness' and 'bias' are related but distinct in the field of education measurement. Mitchell and Cawthon (2014) define bias as 'group-based differences in performance between examinees who are otherwise equally proficient, knowledgeable, skilled, or competent' (p.2). On the other hand, Mitchell and Cawthon (2014) referred to fairness as 'the consequences of test score use'. They argued that 'the performance differences that establish bias, by themselves, do not mean that a test is unfair. Bias only becomes unfairness when the reason for the differences is unrelated to the content of the test' (p. 2).

The 1999 Standards for Educational and Psychological Testing (American Educational Research Association [AERA], American Psychological Association [APA], and National

Council on Measurement in Education [NCME], 1999) provided criteria for test evaluation, testing practices and the effects of test use. The *Standards* vision describes test fairness as follows:

A full consideration of fairness would explore the many functions of testing in relation to its many goals, including the broad goal of achieving equality of opportunity in our society. It would consider the technical properties of tests, the ways test results are reported, and the factors that are validly or erroneously thought to account for patterns of test performance for groups and individuals.

(p. 73)

Wesche, Paribakht and Ready (1996) suggest using multiple measures on a test is preferable to single-format tests. That is, if a single format is used, the test may be biased and in favour of learners who are better at this single method. This also has implications for the reliability of a test. If a test comprises of a variety of formats, it may be more reliable as it is not testing skills through a single method. To reduce this particular test bias, Green (2012) has advised to use different test methods instead of single format.

According to Camilli (2013), ‘individual fairness requires standardised conditions of testing in which standards are treated comparably. This type of fairness is denoted as equity’ (p. 105). Camilli (2013) further illustrated examples of two students with different backgrounds. The first student has some difficulties in learning, such as dyslexia, which could affect his performance on a test, while the other student could be an English language learner. These background-related differences ‘may lead to lower performance under standard testing conditions’ (p. 105). In another words, ‘if a test or test item is equitable, it is presented to individuals under impartial conditions, meaning that no individual student is favoured over another in demonstrating what they know or understand’ (p.105).

Another important component of individual fairness compromises is according to Camilli (2013), ‘treating test takers with dignity and sensitivity. This aspect of testing may have no counterfactual stated in terms of alternative tests or test conditions: It is no defence of a charge of unfairness in this regard to argue that examinees were treated badly but equitably’(p. 106). Sensitivity problems on a test may lead to statistical bias and eventually to faulty interpretation of test scores (Camilli, 2013, p. 107).

A further issue affecting fairness concerns cultural differences and the test takers' familiarity or knowledge of the target culture. Invalid testing due to cultural differences between test maker and test taker is a form of test bias, known as cultural bias (Anastasi, 1976). Bachman and Palmer (1996) argue that one needs to consider the effects of cultural bias in second / foreign language testing as well as testing for cultural bias as part of the evaluation of a test's usefulness.

### **3.19. Teachers' beliefs:**

*Teachers' beliefs* is a term, which is usually employed to describe teachers' pedagogical beliefs, or those of relevance to an individual's teaching (Borg, 2001). The term 'perception' has come to be used as what Bonner (2016) suggested: 'an 'umbrella term' since 'concepts, attitudes, values and beliefs' are similarly shaped by 'mental interpretations'(p.21).

Furthermore, Opre (2015) argues that the preferred term to refer to beliefs about assessment is that of *conceptions*. She argues that this is also the term most frequently employed in the literature. The term *conceptions* draws upon the history of the idea of concept as an abstract object of cognition (Laurence and Margolis, 2012), while the term *teachers' beliefs*, is a term generally used to describe teachers' pedagogic beliefs, or beliefs relevant to an individual's teaching. The most commonly explored areas are teachers' beliefs about teaching, learning, learners and subject matter (e.g. EFL or language) as well as the self as a teacher and the role of a teacher (Calderhead, 1995). Interest in this field of research originates from the premises that conceptions of assessment, as well as teachers' beliefs more generally, 'significantly influence their decisions and professional activity' (Brown, 2008; Vandeyar and Killen, 2007).

In this study, emphasis is placed on teachers' beliefs about their students, the OOPT as a placement test, the modular system and the curriculum structure of the ELI, which are all aspects of the context in which they work. Research on *teachers' beliefs* in EFL has primarily focused on 'teachers' beliefs about teaching, learning and learners; subject matter (i.e. EFL or language); self as a teacher, or the role of a teacher' (Calderhead, 1995, in Borg, 2001, p. 187).

It has been previously pointed out that ‘teachers’ beliefs influence teaching practice and decision-making processes, and ultimately strongly determine the nature of classroom life and the overall classroom atmosphere’ (Borg 2003, 2015; Kalaja et al. 2016). Therefore, understanding teachers’ perspectives can help illustrate how relationships and interactions between teachers and students develop. Research has further suggested that teachers’ practices can inform their belief systems but the two are best understood as mutually informing (e.g. Breen et al. 2001; Burns 1992; Johnson 1992; Li 2013; Phipps and Borg 2009). While it is understandable that beliefs influence practice, teaching practice can likewise inform beliefs. Thus, both teacher beliefs and practices define and shape life in the classroom as well as the opportunities provided to learners. Teachers are central to the learning process and thus their attitudes and beliefs are of vital importance. According to Woodrow (1991) and Levin and Woodman (2006), many teachers are seen to be active agents in the process of suggesting changes and implementation of new ideas and their beliefs and attitudes may either support or impede the success of any educational reform.

### **3.19.1. What shapes teachers’ beliefs?**

Teachers’ beliefs about their students are often shaped by their own experiences initially as learners and later as educators. Borg (1999) argued that there are three key factors that affect language teachers’ beliefs: prior language learning experience, teacher education, and classroom practices. A number of researchers have also emphasised that prior language learning, such as school experience, often impacts teachers’ beliefs about certain aspects of language teaching (Ellis, 2006; Mattheoudakis, 2007; Busch, 2010; Hassan, 2013).

Research shows that teachers’ experiences as learners continue to shape beliefs about teaching and learning throughout their career. Since teacher beliefs and practices are mutually informing, contextual factors are key in determining the extent to which teachers have the freedom to shape the curriculum and the approach adopted towards teaching (Borg, 2003). Furthermore, beliefs may be acquired from peers, teachers or the home environment. These factors may bring teachers to make broad generalisations about their students (e.g. about their ability), which could result in stereotyping or stigmatizing.

Stereotyping refers to the over-generalisation of characteristics applied to group members without consideration of the differences within that individual or group (Chang and Sue, 2003;

Reyna, 2000). Stigma, on the other hand, refers to the devaluation of overgeneralisations made about a group (Dovidio, Major, and Crocker, 2000; Goffman, 1963). Stereotypes or stigma can result in unfair assumptions or misattributions of the values, behaviours, and interests of group members perceived as different by dominant society and, because it is common to assume that group members share similar characteristics, it is more likely that a stigmatised group be affected by the negative consequences of stereotyping compared to their non-stigmatised peers (Biernat and Dovidio, 2000). Tenenbaum and Ruck (2007) suggested that teachers may also hold stereotypes about other areas aside from academic ability. For instance, they may make assumptions on the language proficiency of students, social competency or work ethic. These stereotypes may ‘negatively influence teachers’ placement decisions’ (Tenenbaum and Ruck, 2007 in Riley, 2015, p. 663). Understanding the dimensionality of teacher conceptions about assessment may help us better understand ‘questions about the effects of training and policy directives on outcomes’ (Bonner, 2016, p.26).

### **3.20. What affects students’ test performance?**

A number of factors may affect students’ performance on tests; these include affective and educational factors as well as other factors. Foreign language researchers have long been aware that language learning may be associated with affective factors, such as anxiety and motivation, which in turn may affect performance (Mohammed, 2016). Furthermore, there are factors related to the student’s previous educational experience, which can also influence test performance.

O’Sullivan and Green (2011) suggest that there are three kinds of test taker characteristics, which can influence test performance: 1) Physical/physiological, 2) Psychological and 3) Experiential. These are outlined in the figure below.

**Figure 7. Characteristics of assesses - O’Sullivan and Green (2011)**

<b>Physical/Physiological</b>	<b>Psychological</b>	<b>Experiential</b>
	<b>Unpredictable</b>	<b>Short term</b>
<b>Short term illness / injury</b>		<b>Examination preparedness</b>
	<b>Concentration</b>	<b>Examination experience</b>
	<b>Emotional response</b>	<b>Communication experience</b>
<b>Age</b>	<b>Motivation</b>	
	<b>Memory resources</b>	<b>TL-country residence</b>
	<b>Cognitive style</b>	<b>Topic knowledge</b>
<b>Long-term disability</b>	<b>Personality</b>	<b>Knowledge of the world</b>
		<b>Education</b>
<b>Gender</b>	<b>Systematic</b>	<b>Long term</b>
		<b>Language background</b>

*Adapted from Green (2014, p.69)*

O’Sullivan and Green (2011) suggest that these characteristics can influence performance on assessments, for example some students do very well under test conditions, while others suffer from ‘debilitating anxiety’ and may be ‘unable to participate’ as a result of their psychological characteristics (Green, 2014, p.68). Similarly, physical characteristics such as illness and disability may impact on test performance. Experiential characteristics such as previous practice with certain types of assessments or general knowledge can also impact on performance (ibid.).

Background factors worthy of investigation include age, gender, primary language background (L1), and level of education or subject specialism in the case of applicants for university places (Fulcher, 1999). Alrabai (2016) identifies a number of student-related factors that can play a role in achievement for Saudi EFL students.

Low English language achievement can be attributed to a variety of interrelated factors ranging from learner-related variables (e.g., gender, motivation, and anxiety) to sociocultural variables (e.g., the influence of L1, society, culture, and religion). In addition, the variables of EFL instruction (i.e., teacher behavior and practices, the curriculum, and the teaching method) and other factors related to problems with the educational system in Saudi Arabia (e.g., overcrowded classes, a lack of teacher training, and a lack of technology) may also contribute to the poor EFL results.

As Alrabai (2016) illustrates, that there are a number of factors affecting performance, some of which are more closely related to the background of students, while others are related to teachers and aspects of the curriculum.

### **3.20.1. Psychological factors**

Psychological factors, as described in O'Sullivan and Green's model (2011), mentioned above, include characteristics such as motivation, emotional response and personality. Among these, affective factors can influence students' attitudes towards learning and performance. Test-anxiety, for instance, has been found to have detrimental effects on students' psychological well-being (Zeidner, 1998; 2014).

Affective factors may be understood as emotional factors, which influence learning. These can have both positive and negative effects; negative affective factors are referred to as 'affective filters', which draw on the *Affective filter hypothesis* first proposed by Dulay and Burt (1977), which was later incorporated in Krashen's (1985) *Input hypothesis*. Krashen (1985) argued in order for learners to be able to acquire a second language, they need to obtain 'comprehensible input' and the 'affective filters' need to be low enough to allow the learner to process the information. According to his theory, affect includes motivation, attitude, anxiety and self-confidence. While there have been some criticisms to the input hypothesis, the terminology of 'affective factors' is widely used in Second Language Acquisition (SLA) research. Affective factors, more generally, are seen to play a significant role in SLA. According to Krashen (1985), when the 'affective filter' is high, this can create a barrier to acquisition. The affective filter is high when the learner is either unmotivated, lacks in confidence or may be concerned with failure and thus experience great anxiety. The affective filter is low when the learner is not anxious and is motivated to learn. As mentioned above, Krashen (1985) theorised four factors, which can influence SLA. These are the following: motivation, attitude, anxiety, and self-confidence. Further research into SLA has identified these factors, among others, as contributing to SLA.

### 3.20.1.1 Motivation to learn English

The word motivation comes from the Latin *movere*, which means ‘to move’. Motivation, therefore refers to ‘what moves a person to make certain choices, to engage in action, to expand effort and to persist in action’ (Dörnyei and Ushioda, 2013, p. 3).

Motivation has been identified by researchers and educators as a key factor for successful language learning. Without sufficient motivation, even students with the most remarkable abilities struggle to accomplish long-term goals, and even appropriate curricula and good teaching are not enough to ensure achievement (Dörnyei and Csizér, 1998).

Dörnyei and Ushioda (2013) make reference to the well-known distinction between *intrinsic* and *extrinsic* motivation. *Intrinsic motivation* (IM) refers to behaviour exhibited in order to experience pleasure and satisfaction, such as the joy of doing an activity or satisfying one’s curiosity. *Extrinsic motivation* (EM), on the other hand, refers to performing a behaviour as a means to an end, for instance to receive a reward or avoiding punishment. Dörnyei and Ushioda (2013, p.23) also mention a third term, i.e. *amotivation* (AM), which refers to the lack of any kind of motivation, whether intrinsic or extrinsic. This distinction was first made within psychological theory and later become employed more broadly. Bandura (1997) for instance, describes intrinsic motivation as the enjoyment or love for a particular activity. Traditionally, extrinsic motivation has been viewed as something, which can undermine intrinsic motivation. Lepper and Greene (1978), for instance, found that students lost their natural intrinsic interest in an activity if they had to meet some extrinsic requirements. Deci and Ryan (1985) and Ryan and Deci (2000), however, criticised this dichotomy and suggested that replacing the distinction between extrinsic and intrinsic motivation with the principles of what they termed *self-determination theory* (SDT). In their view, extrinsic goals may be ‘fully internalised within a person’s self-concept’ (Dörnyei and Ushioda, 2013, p.24). For example, a learner may value being able to speak a particular language but also enjoy the learning process. Deci and Ryan (2000) thus, view extrinsic motivation as behaviour, which can either be entirely externally regulated (motivated by external rewards or fear of punishment), or in part externally regulated (through self-control in order to comply with expected norms) and also somewhat internally motivated (when people attribute personal importance to an activity).

There are a number of factors that can contribute to students’ motivation for studying a second language. Ushioda (1996) points out that cooperative learning can increase intrinsic



motivation. For some students, therefore, working in groups, rather than individually, is preferable as it provides them opportunities to work together and share ideas or experiences with one another. This experience can also increase the potential for superior learning. Some studies on motivation have suggested that the familiarity of the content, as well as opportunities to compare the foreign culture to their own, may increase intrinsic motivation for some students. A study carried out in Chile found that learners developed their intrinsic motivation when they were presented with a variety of different English language contexts, which they could compare to their own culture (Glas and Cardenas-Claros, 2013). The authors suggest that learners need to be exposed to a combination of both global and local contexts so that they understand the cultural context within which English is employed as the lingua franca. This combination can also help students understand how English is used within their own cultural context.

Students studying a second language can also be significantly influenced by the social and cultural context. Teachers, for instance, can affect students' motivation in both positive and negative ways. Dörnyei and Ushioda (2013) place emphasis on teachers' perspectives on students' motivation yet highlight that the role teachers play in engaging students in the learning process is complex as 'it concerns almost all academic and social aspects of the classroom environment' (Kubanyiova, 2006 cited in Dörnyei and Ushioda, 2013, p.28).

Veenman (1984), examined the perception of the problems that teachers new to the profession face, found that teachers ranked *motivating pupils* as the second most serious source of difficulty. While Veenman (1984) explored the views of novice teachers, 'the question of how student motivation can be increased remains a prevailing issue for seasoned practitioners as well' (Dörnyei and Ushioda, 2013, p.105). Teachers' beliefs regarding student motivation can influence the efforts expended and strategies used for motivating students (Hardré et al., 2006). It is thus of paramount importance to understand teachers' beliefs regarding the goals that students have, their interest and perceived value of the material, and teachers' beliefs regarding the nature of student motivation (Hardré and Sullivan, 2008, 2009; Kaplan and Maehr, 2007).

Furthermore, social milieu, comprising of the teachers, students and the classroom climate, which develops as a result of the interaction, can also influence motivation. Research has

repeatedly demonstrated the relationship between classroom environment and student motivation (Greene et al. 2004; Hardré et al. 2007; Hardré and Sullivan, 2008). Furthermore, teachers' supportiveness as well as their interpersonal style, has also been found to influence student motivation (Anderman and Wolters, 2006). Teachers' perceptions and beliefs can thus influence both classroom practice and the effort teachers make to increase student motivation (Linnenbrink and Pintrich, 2002).

With regards to the Saudi context, Alrabai (2014) argued that teachers generally focus more on their teaching practices and completely neglect the motivational aspects of the curriculum. Alrabai (2014) further reveals that the EFL curriculum in Saudi Arabia does not address 'students' interests, needs, goals, experiences, daily life activities and real-world situations (pp. 242-243). The overwhelmingly teacher-focused approach adopted in this context can further affect students' motivation as they may feel that the topics are not relevant to them and that their perspective on the learning experience is not valued. Al-Seghayer (2011) claimed that the key missing ingredient for most Saudi students is intrinsic motivation, which is essential for becoming proficient users of English. Saudi students learn English because it is an educational requirement, hence as a result of external pressures, yet they have minimal internal motivation and their attitude towards learning English are often negative (Al-Seghayer, 2017, p.39). Similarly, Mahmoud (2014) also found that Saudi students frequently lack motivation towards studying English, which can affect the effort that students make towards their learning.

Aside from the previously mentioned factors, there are also internal factors that might influence Saudi students' lack of motivation for learning English. Al-Rabai (2014), for instance, found that low self-esteem, self-confidence, and motivational intensity strongly influenced Saudi EFL students' learning. Elyas and Al-Grigri (2014), who explored teachers' perspective on the motivation of Saudi students for learning English, found that students exhibited low levels of motivation for the subject. As Ahmad (2015) suggests, some Saudis hardly acknowledge that English is an important skill for success in educational and occupational careers. This has implications both for students' motivation while studying the language and for the educational outcomes.

### **3.20.1.2. Anxiety and stress**

Language anxiety and stress can affect the way learners perform both on tests, such as the OOPT, as well as on a language course. Speaking, in particular, is often a greater cause of stress to students compared to the other skills of reading, listening and writing, as other students as well as the teacher can evaluate the performance. The extent to which a student is more or less familiar with a specific task can have an impact on the degree to which they feel unsure or uncomfortable with the exercise or assessment. Test-related anxiety can also affect the results students obtain on a test. The measurement properties of an achievement test could, for instance, differ across students with low, medium and high test-anxiety. This means that it would not be possible to compare test scores across these groups of students (Vogl and Pekrun, 2016).

Numerous studies carried out in Saudi Arabia have found that language anxiety manifests itself prominently among most Saudi EFL learners (Alrabai, 2015; Al-Saraj, 2014; Javid, 2014; Mohammed, 2016). Alrabai (2014) stressed that, in some instances, Saudi students' low competence in English could be linked to anxiety. He argued that students often appear reluctant to take part in classroom discussions and hesitant in providing responses when asked. They are generally unwilling to engage in speaking activities and heavily rely on the teacher. This was found to be a concern for both male and female students, who were found to exhibit the same degree of language anxiety (Alshahrani and Alandal, 2015). The extent to which a student feels uneasy communicating in the classroom is likely to be linked to the teaching methods adopted in both Saudi schools and HE. Aljumah (2011) asserts that students are offered insufficient opportunities to practise the TL and therefore find it difficult to speak in class and, when they do speak, these are brief utterances. He added that in the Saudi EFL classroom 'teachers and students are usually on opposing sides: teachers speak constantly, whereas students mumble and swallow their words or say nothing (p. 85)'. Rabab'ah (2002) correspondingly emphasised that, both in public schools and at university, students in Saudi Arabia are not provided with sufficient opportunities to practise speaking English in the classroom.

### **3.20.1.3. Experiential characteristics**

Experiential characteristics include a variety of factors such as examination preparedness, previous education, language background, topic knowledge as well as attitude towards language. Attitudes towards studying English may be understood in two ways. On the one hand, one can reflect upon how students view English as a subject and hence look at the attitude they have towards the learning activities, content of the lessons and tests, as they have experienced them. On the other hand, one can look at students' attitudes towards learning a language, which is spoken across the world (as an international language) and as a home language in English-speaking countries. In recent years, there has been a shift in attitudes towards studying English. For instance, Faruk (2014) stated that Saudis' attitudes toward English are becoming increasingly positive, as English has become vital to the country's future prosperity and English language skills are now required in many fields. Similarly, Alrahaili (2013) found that Saudi EFL learners hold generally positive attitudes toward English and its speakers. They are, however, generally opposed to accepting the TL group's social and religious values. This is understandable as a result of the significant cultural differences and the extent to which religion in Saudi Arabia influences aspects of daily life. Other studies considered the positive shift in attitude has recently become more inclusive of the culture of the target culture. Hagler (2014) found that a clear majority of students had a positive disposition toward Western culture and stated that they liked some aspects of it. However, attitudes towards learning English may be less positive as a result of dissatisfaction with a previous learning experience (e.g. from schooling) or from a student's current learning experience at university. Students' attitude towards learning English is strongly interconnected with motivation, and therefore it is important that teachers are aware of the interests that students may have for the subject area, so that they can make the curriculum more relevant. As Dörnyei (2001) suggests, relating the subject content to students' everyday experiences and backgrounds can increase student motivation.

Students' prior learning experiences as well as socio-economic status (SES) can also affect test performance. According to Parson, Henson and Sardo-Brown (2001), 'SES is the term used to distinguish between people's relative position in society in terms of family income, educational background and occupational prestige' (cited in Ariani and Ghafournia, 2015, p.17). Higher SES can result in greater opportunities with regards to education and travel,

which may result in different learning experiences. With reference to this study, this is also relevant to students' language background as there are inherent differences in the way English is taught in private and public schools within Saudi Arabia.

### **3.20.2. Physical and physiological characteristics**

There are also physical characteristics that may influence a test taker's performance. These may include factors such as temporary illness, injury or disability. A student may underperform on a test because of health-related factors, which may result in a score, which does not portray an accurate picture of the students' knowledge. Green (2014) argues that performance on tasks of all kinds 'is affected by alertness, health and mood' (p.67). Therefore, 'if we were to administer identical and perfectly accurate language assessments on different occasions [...] we would still see differences in the learners' scores on each occasion' (Green, 2014, p.67). The extent to which health-related factors affect performance can vary significantly between individuals. One way to address this, is to offer students an opportunity to re-sit an examination if they can provide evidence of extenuating circumstances. With regards to long-term disabilities, universities often make special arrangements to ensure students with disabilities are not disadvantaged.

### **3.21. Summary**

The chapter reviewed the literature on testing and placement tests within this broader context, as well as exploring previous studies and theory on teachers' beliefs about what factors may affect student achievement. The literature highlights the importance of understanding the test purposes of placement tests as well as acknowledging the range of factors, which can affect test-takers' performance. This is supported both by a theoretical perspective and by previous empirical research. With regards to the theoretical orientation, the chapter has drawn from authors with a background in language testing, such as Peter Tymms, Anthony Green, Cyril J. Weir, Christine Coombe, education or anthropology in education such as Paul

Gibbs, Bryan Maddox and Caroline Gipps and from psycholinguistics, such as Zoltan Dörnyei.

The chapter has also drawn on empirical studies, which focused specifically on employing the OOPT, such as Sinlapachai et al. (2016), Mahmoud (2014) and Brooke et al. (2012).

The number of studies, which explored the use of the OOPT in Saudi Arabia or similar contexts, however, is minimal, therefore, this study makes a significant contribution to the current empirical research in this field. The chapter further describes the value of teachers' beliefs, which is central to the investigation, since what is learnt about students is portrayed through the eyes of their teachers. The literature review, however, also highlights the lack of studies on how cultural factors may shape these beliefs. This study thus, adds to the existing literature by exploring the role of culture in the development of teachers' beliefs and conceptions with regards to placement tests such as the OOPT. It has been noted that, in Saudi Arabia, students make slower progress than anticipated in English, however there has not been sufficient research on the reasons for this.

## **CHAPTER 4 RESEARCH METHODOLOGY**

### **4.1. Introduction and research aims:**

This chapter describes the theoretical assumptions behind the research design and the methodology adopted in this study. It begins by outlining the research questions, followed by a discussion on the aims, participants sampling, data collection, analysis validity, reliability and ethical considerations. The chapter then reviews the differences between various paradigms followed by a discussion of the research paradigm adopted for the study. It further defines ‘case study’ as a research approach and outlines the benefits and challenges of mixed-methods research. The chapter also describes the participants and sampling method and the stages of the data collection and analysis process.

This research project aims to investigate the use of the OOPT in an EFL context, namely within the ELI at KAU in Saudi Arabia. It is important to note, that this study does not intend to test pre-existing theories or hypotheses, but rather seeks to explore teachers’ beliefs about the OOPT and students in the ELI. It is hence guided by an inductive rather than deductive approach. The rationale for exploring this particular phenomenon emerged from my own experience as an ELI teacher. Colleagues in the ELI often expressed resentment towards the OOPT, and it was apparent that they held strong views about the students taking the placement test at KAU. Some of these concerns prompted me to investigate teachers’ views on the OOPT in greater depth.

As previously mentioned, the OOPT was designed to measure students’ English language ability in order to place them into an appropriate English language course. At KAU, the OOPT is used for this purpose and students are placed into one of the four levels of proficiency offered by the ELI on the basis of their OOPT scores. A breakdown of the scores and their corresponding levels is outlined in chapter 2. The study employed both qualitative and quantitative data collection methods. The latter increased the number of participants, thus making the findings more reliable, while also including rich qualitative data, which could help gain a more in-depth understanding of reasons why teachers held certain views of the test and their students. The rationale for employing both a questionnaire and interviews

is further discussed in this chapter in relation to the research design. The research instruments adopted consisted of a questionnaire composed of both closed and open questions and semi-structured interviews with teachers in the ELI. The context of the data collection is of great importance to the study, as I chose to explore the phenomenon within the department where I work and hence, I approached the investigation from an insider's perspective.

#### **4.2. Research questions:**

Although adopting a placement test may facilitate the placement of students into an appropriate English language course, in my view, having taught at the ELI myself, the OOPT did not appear to be as effective as it could be for students and teachers at KAU. While numerous studies have explored issues relating to placement tests, very little is known about the challenges of employing a placement test in this specific context. This study aims to fill this gap in the literature in order to provide a better understanding of teachers' perspectives on the OOPT, on the ELI modules and, more generally, on students taking the test within the ELI.

It has been suggested that engaging with the literature in the early stages is not advisable as it may lead the researcher to place greater focus on certain aspects of the data at the expense of other potentially important ones (Braun and Clarke, 2006). However, in my view, my early reading of the literature informed the writing of the research questions and helped increase my engagement with the data.

This study is guided by four objectives (see section 1.3), from which the following research questions are drawn:

- 1. How do teachers view the OOPT?**
- 2. How does the OOPT impact on placement and teaching, in the opinions of teachers?**
- 3. What are the factors that might affect students' achievement in both the OOPT and during the English module, in the opinion of teachers?**
- 4. How can the system be improved to support students' achievement?**



The first and fourth research questions drew primarily from my own teaching experience in the ELI and the informal discussions I had with some of my colleagues, which identifies some issues with the OOPT and the ELI modules. The second research question, on the other hand, was informed by the literature on placement tests and on teachers' beliefs. Similarly, the third research question drew from the literature on factors affecting students' achievement, in particular the work of O'Sullivan and Green (2011) (see section 3.16).

The first three research questions rely both on the qualitative and the quantitative data, while the fourth question relied predominantly on the interview data. As this study is exploratory in nature, it is primarily concerned with interpreting realities through the lens of the participants and thus the aim of the investigation is to better understand a phenomenon (i.e. teachers' beliefs on the OOPT and their students). The naturalist paradigm (see Lincoln and Guba, 1985) best describes the approach towards the data collection and analysis, in particular because of the strong emphasis on teachers' beliefs. Thus, comments on the OOPT and students sitting the placement test in the ELI need to be interpreted through the perspective of teachers.

Research questions 1 and 2 also seek to collect data about the validity of the OOPT, understood as a unified view of validity (Messick, 1989), while questions 3 and 4 seek to gain a better understanding of factors which, according to teachers may affect test performance as well as ways in which the current ELI programme may be improved.

#### **4.3. Mixed-methods research:**

The division between qualitative and quantitative research has historically been represented as 'unbridgeable'. Exponents of each approach 'argued that theirs was the only true way' to undertake research in their field and 'the debates between them became so argumentative that the phrase 'paradigm wars' was used to describe them'. (Newby, 2014, p. 96)

The paradigm wars (Gage, 1989), in which researchers stood by their allegiances either to quantitative or qualitative methodologies, have given way to mixed-methods research (Gorard and Taylor, 2004; Gorard and Smith, 2006; Teddlie and Tashakkori, 2009). Denzin,

(2008, p.322) notes that there is a need for less confrontational approaches to be adopted between different research paradigms. There is also a need for greater convergence between the two (Brannen, 2005) and greater dialogue between their proponents. There has been much criticism against the polarisation of research into either qualitative or quantitative approaches and their associated objectivity and subjectivity respectively (Ercikan and Roth, 2006). Onwuegbuzie and Leech (2005, p.382) suggest that the terms 'qualitative' and 'quantitative' should be replaced with 'confirmatory and explanatory research'. While these terms better describe the aims of qualitative and quantitative research, they still present a dichotomy between the two approaches.

Mixed methods research, on the other hand, distances itself from these polarities and argues for their compatibility (Denzin, 2008; Trifonas, 2009, Creswell, 2009). The concept of mixed-methods research, also referred to as triangulation, was introduced by Campbell and Fiske (1959). It refers to an approach that uses 'multiple observers, theoretical perspectives, sources of data, and methodologies' (Denzin, 1970, p. 310). Tashakkori and Teddlie (2003) urged researchers to recognise and use the term 'mixed methods' more precisely and 'the use of 'mixed' as an umbrella term to cover the multifaceted procedures of combining, integrating and inking the different types of methods and data' (Punch, 2011, p. 289).

Some researchers have considered the mixed-methods approach to be a third methodological movement in educational and social research (Johnson, Onwuegbuzie, Turner, 2007; Tashakkori and Teddlie, 2003). The main advantage of gathering qualitative data in this study is that it allows the researcher to explore a specific phenomenon in greater depth. Denzin (1989) suggested that qualitative approaches produce thick (detailed) description of feelings, opinions and experiences. Quantitative data, on the other hand, can address some of the issues inherent in qualitative approaches, such as small sample sizes. A further advantage of combining quantitative and qualitative approaches is that the data may expose differing perspectives as well as verify findings so that an issue can be examined in more detail (Bryman, 2004).

In similar vein, Creswell and Clark (2007) claim that data should be mixed together in order to form a complete picture of the problem, rather than simply collecting and analysing quantitative or qualitative data. Creswell and Clark (2007, p. 9) argued that mixed methods

research ‘provides strengths that offset the weaknesses of both quantitative and qualitative research’. For instance, it has been suggested that quantitative research is weak in understanding the research context or setting. In this manner, mixed-methods research can provide ‘more comprehensive evidence for studying a research problem than either quantitative or qualitative research alone’ (ibid.). The following sections outline the research design and method adopted in the investigation.

#### **4.4. Methodological approach: case study:**

The study adopts a case study approach comprising of a questionnaire and semi-structured interviews. The uniqueness of the socio-cultural context is of key importance to the investigation, hence a case study appeared to be the best approach to fully capture some of the complexities related to this learning environment.

There are various definitions of what constitutes a case study yet different researchers have defined the method according to their own perspectives. Case studies generally seek ‘to understand the case in depth, and in its natural setting, recognizing its complexity and its context’ (Punch, 2012, p. 144). They can be useful for providing detailed descriptions and for portraying to the readers ‘what it is like to be involved in this situation’ (Lincoln and Guba, 1985, p. 214). A case study can be described as ‘an inquiry that investigates a contemporary phenomenon within its real-life context’ (Yin, 1994, p. 13). Furthermore, case studies enable researchers to explore a specific case over time, ‘through detailed, in-depth data collection involving multiple sources of information [...] and reports a case description and case-based themes’ (Creswell, 2007, p. 73).

In the field of Teaching English to Speakers of Other Languages (TESOL), the term ‘case study’ refers to the analysis of ‘a person, either a learner or a teacher, or an entity, such as a school, a university, a classroom, or a program’ (Chapelle and Duff, 2003, p. 164). Stake (2005) reveals that case studies are not a methodology, but rather a choice of what is to be studied within a bounded system. On the other hand, (Denzin and Lincoln, 2005; Merriam, 1998; Yin, 2003) consider a case study as a methodology or research strategy. This involves gathering multiple sources of evidence (such as interviews, questionnaires and observations) in order to report on case-based themes.

Yin (2003) categorises case studies as *explanatory*, *exploratory* or *descriptive*. An exploratory case study appeared best suited for my investigation as the study seeks to find out how teachers view the OOPT, the ELI modules and students within the ELI. Among the factors, which supported the decision to adopt a case study design, are the following:

While there are a number of advantages of conducting case study research, ‘a common concern of case studies is that they lack generalisability’ (Punch, 2005, p.145). If a study is based on one case alone, it cannot be generalised to a wider population. There are two types of case study situations in which generalisation would not be the objective:

First, the case may be so important, interesting, or misunderstood, that it deserves a study in its own right. Or it may be unique in some very important respects, and therefore worthy of studying’

(Punch, 2005, p.146)

Both situations are relevant to the present study, as the use of placement tests within this specific context is both unique and under-researched and thus, it could be argued that it deserves a study in its own right.

#### **4.5. Research design:**

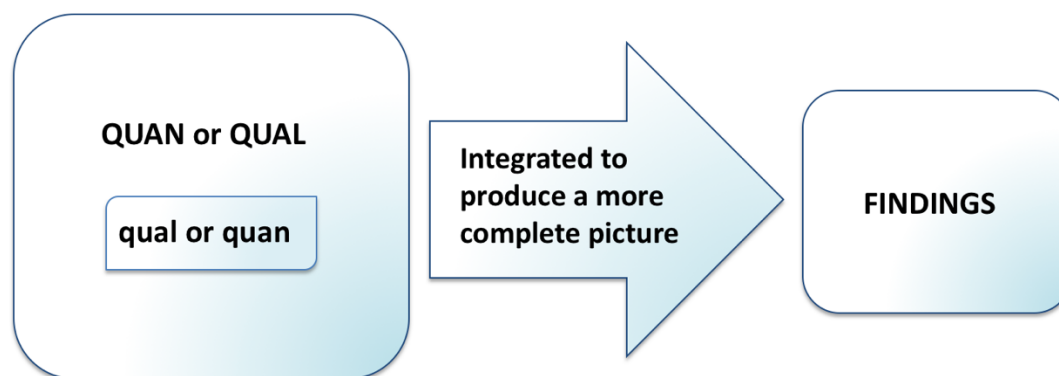
The research design comprises two interviews and a questionnaire and thus involves both the collection of qualitative and quantitative data. Collecting both types of data enables the researcher to seek answers to the same research questions but using different kinds of methods by ‘gathering and interpreting data from multiple viewpoints’ (Brown, 2014, p. 20). With regards to exploring human behaviour specifically, Cohen et al. (2007) claim that mixed-methods research can ‘map out, or explain more fully, the richness and complexity of human behaviour by studying it from more than one standpoint and, in so doing, by making use of both quantitative and qualitative data (Cohen et al., 2007, p. 141). With reference to this study, mixing methods was found to be best suited as it enables the researcher to collect responses from a wider population, while also capturing rich qualitative data from the interviews.

According to Bryman (2001), ‘a research design provides a framework for the collection and analysis of data. A choice of research design reflects decisions about the priority being given to a range of dimensions of the research process’ (p.29). In mixed-methods designs, data

can be either collected at the same time (i.e., concurrently) or introduced in phases over a period of time (i.e., sequentially). When the data are introduced in phases, either the qualitative or the quantitative data may be gathered first, but the sequence relates to the objectives being sought by the researcher in the mixed methods study. If qualitative data collection precedes the quantitative data collection, the purpose is to first explore the problem under study and then follow up with quantitative data in order to extend the investigation to a large sample. Alternatively, if quantitative data is collected prior to qualitative data, the aim is to explore with a large sample first and then to follow up in greater depth with a smaller number of cases during the qualitative phase. When gathering both forms of data concurrently, the researcher seeks to compare both forms of data to identify for congruent findings (e.g., how the themes identified in the qualitative data collection compare with the statistical results in the quantitative analysis) (Creswell et al., 2003).

Creswell and Clark (2011) propose a number of mixed-methods designs, which are helpful to illustrate the different ways in which researchers can combine both qualitative and quantitative data. The design adopted in this study is best described as an *embedded design*. The figure below illustrates the *embedded* model as theorised in Creswell and Clark (2011).

**Figure 8. Embedded design Creswell and Clark (2011)**



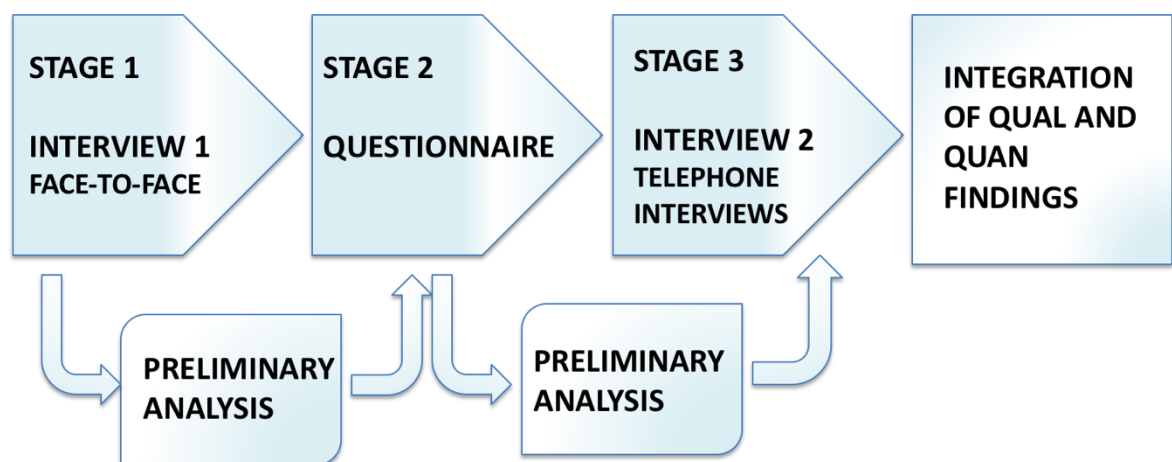
Bryman (2016) proposes that ‘the need for an embedded design can arise when the researcher needs to enhance either quantitative or qualitative research with the other approach’ (p. 640). He further added that in embedded designs ‘the phasing of the data may be simultaneous or sequential’ (ibid.) and suggests that *embedded designs* are appropriate where ‘the researcher

feels that quantitative (or qualitative) research alone will be insufficient for understanding the phenomenon of interest' (ibid.).

An analysis of each of the three data sets was carried out after its collection and this initial interpretation informed the subsequent phase. However, data from all three phases were also analysed in greater depth once all the data was collected and findings from both qualitative and quantitative data were checked against each other.

Unlike Creswell et al., (2003), who suggest that the collection of one type of data precede the other in sequential designs, Brown (2014) argues that each of the data collection methods feeds into the other and that multiple repetitions of these patterns could be productive (p. 47). For example, qualitative sampling could be used to discover and formulate the issues, questions, hypothesis etc. for a study. This may be followed by quantitative sampling, that would investigate those issues, questions, hypotheses, etc. on a large scale followed further by qualitative sampling to help the researcher understand the quantitative results in more depth and so forth (ibid., p. 47-48). The approach towards the data collection in this study mirrors the multiple repetitions of sequencing qualitative and quantitative sampling described above. The figure below illustrates how a preliminary analysis informed the subsequent stages of the data collection as well as how both qualitative and quantitative findings were integrated once the data collection was complete.

**Figure 9. Research design**



While the interview data takes precedence over the questionnaire data, as the research design comprised two interview stages, the quantitative data provides a much broader picture.

The first interview was informed by the research questions in addition to the initial literature review and it consisted of face-to-face interviews with teachers of the ELI. Following a preliminary analysis of the Stage 1 interviews, the questionnaire was designed and handed out to participants. The questionnaire results were then analysed statistically and over a year later telephone interviews were conducted with a different set of participants within the ELI. The second interview stage was hence informed by the first data collection stage. Questions in the interview schedule (for Stage 2 interviews) were modified as a result of the preliminary analysis of the Stage 1 data sets.

The initial analysis of the face-to-face interviews also exposed how teachers appeared to equate the concept of validity with the concept of effectiveness. In the stage 1 interviews, teachers were asked whether they felt the OOPT was valid, however their responses described its perceived effectiveness rather than its validity. For instance, teachers referred to the OOPT as generally effectively placing students in an appropriate level. Where teachers made criticisms of the OOPT, for instance with reference to its cultural content and how it may disadvantage some students, this was not linked to test validity. It thus emerged that the concept of effectiveness better encompassed the views teachers expressed on the OOPT. As a result greater emphasis was placed on the concept of effectiveness rather than validity. The following section describes the data collection, research population and sampling method adopted in the study.

#### **4.6. Data collection, research population and sampling:**

The participants in this study were all teachers from the ELI department (female section) at KAU. There was an initial interview (Stage 1) in which eleven teachers volunteered to take part. For convenience, the questionnaire (Stage 2) was handed out at the same time of the interview although not all teachers interviewed returned the questionnaire. One hundred questionnaires were handed out of which sixty were returned. The questionnaire results informed the second set of interviews (Stage 3), which was carried out remotely through

telephone interviews (see appendix 5). However, not all teachers who were interviewed in the second interview stage completed the questionnaire. Seven teachers agreed to be interviewed in the second stage.

The participants were heterogeneous in respect of nationality, mother tongue and teaching experience. Some of the newly appointed teachers, all non-Saudi citizens, refused to complete the questionnaire because they were unsure about how to respond to questions, which related to specific components of the curriculum (such as the OOPT). Some others were hesitant and refused to take part without providing clear reasons. In both cases, these teachers were unfamiliar with the researcher. It was evident that teachers who had been working at the ELI for longer periods of time, and had known the researcher as a colleague, generally took part in the study.

Purposive sampling was used in this study as participants were all teachers of the ELI, and were therefore purposively selected, however participants chose whether or not to complete the questionnaire or take part in the interview. The questionnaire was handed out to the majority of teachers present at the ELI at the time of the data collection, but it was not possible to include all teaching staff because of the time constraints and teachers' availability. Sampling which does not include the whole population can be regarded as an acceptable alternative if the research population is large or if the cost or time associated with data collection is high (Bryman and Bell, 2003). The research opted for the use of purposeful non-probability sampling (Bryman, 2008; Creswell, 2014) and teachers were selected according to two criteria: purposiveness and accessibility (Silverman, 2001). As the study aimed to explore teachers' views on the OOPT within this specific context, there was a purposively selected target population, i.e. teaching staff within the ELI.

In line with naturalistic inquiries, understanding individual experiences in a natural setting remains the focus of the investigation. Thus, purposive sampling is the most appropriate sampling method for accomplishing this aim (Lincoln and Guba, 1985; Robson, 2011). Purposeful or purposive sampling requires selecting participants on the basis of specific criteria (Brown, 2014). Participants are often purposively selected because they can provide useful and valuable information about the research topic (Denscombe, 2014; Ritchie et al.,



2014). However, a possible drawback of purposive sampling concerns generalisation to a population (Bryman, 2012). This may be viewed as a limitation of the chosen approach, although as the study aimed to explore a phenomenon within a specific context, it was essential to select participants who were employees of the ELI, as they would have the necessary knowledge and experience to provide useful responses.

#### **4.7. Positionality and insider research:**

Conducting research as an insider, means the researcher is already familiar with the context and individuals within this context. Insider research is defined as research carried out within a community by a member of the said community (Brannick and Coghlan, 2007). As an insider, the researcher is in ‘a unique position to study a particular issue in depth and with special knowledge about that issue’ (Costley et al., 2013, p. 3). Insiders also have ‘easy access to people and information’ and are in ‘a prime position to investigate and make changes to a practice situation’ (ibid.).

As an employee of the ELI, I was already informed about the curriculum, current practice regarding placement tests (the use of the OOPT) and some of the issues regarding its implementation. This was certainly an advantage, as it was not necessary to spend time familiarising myself with the research context. However, as Sikes (2006, p. 112) suggests, individuals who choose to carry out insider research ‘have to think very carefully about what taking on the role and identity of researcher can mean and involve in a setting where they are normally seen as someone else with particular responsibilities and powers’. In this study, extra care was taken to ensure the questionnaires and interviews were administered professionally and that participants were fully informed about the aims of the investigation as well as the potential benefits for the ELI.

Among the many researcher who explored advantages and drawbacks of insider research, Bonner and Tolhurst (2002) further point out some of the advantages of insider research, which include greater cultural understanding, decreased impact on the flow of social interaction and participants’ increased willingness to share personal insights and information.

However, the role as an insider may conflict with the research role (DeLyser, 2001, Gerrish, 1997; Unluer, 2012) and familiarity with the context may lead to loss of objectivity e.g. routine behaviours might be overlooked (DeLyser, 2001; Hewitt-Taylor, 2002; Unluer, 2012). There is, therefore, a need for strict ethical compliance as the researcher may have access to sensitive information, which outsider researchers would generally not be able to access (Smyth and Holian, 2008; Unluer, 2012).

It has been suggested that it is impossible to conduct research with no positionality (Punch, 2009) and that insider research is not objective enough as there may be too much emotional input on behalf of the researcher (Alvesson, 2003; Anderson and Herr, 1999; Anderson et al., 1994). However, insiders may have greater insight into the experiences of the participants, which, on the one hand, may be seen as a limitation, but on the other, could also be seen as a benefit. With reference to my study, as an insider I could easily relate to some of the challenges experienced by teachers, for instance with regards to the weak English proficiency of many incoming students or the view that Science students generally perform better than Art students in the OOPT and ELI modules. One of the risks of insider research, however, is that certain assumptions remain unchallenged by the researcher. A possible drawback of being an insider might therefore include a lack of objectivity and a tendency of the researcher to make assumptions based on familiarity (Delyser, 2001; Hewitt-taylor, 2002). To address this, I sought the second opinion of a fellow Phd student, who was not employed by the ELI, on my data analysis.

There are also ethical issues arising from practitioner research that need to be taken into consideration. Practitioner researchers 'work with their colleagues, who are temporarily transformed into research subjects, and then return to colleagues in much the same way as the researcher adopts an academic researcher role and then returns to that of colleague' (Costley and Gibbs, 2006, p. 89). The authors argue that when conducting practitioner research there are 'connections between self and the communities of practice' within the context in which the researcher works. As a result, 'researchers need to become reconciled with these larger structures, and consider any conflict' (p.92).

While the challenges of conducting insider research were this taken into consideration and participants were fully informed about the study (an informed consent, see appendix 2, was signed before the interviews took place), it is understandable that some of the participants may have been influenced by my personal involvement with the ELI as a member of staff. However, my positionality as an insider facilitated both the data collection and analysis, as participants were already familiar with me as a colleague. For most, this increased their wish to participate and sparked curiosity for my investigation.

#### **4.8. Deciding on using synchronous or an asynchronous data collection:**

There is an increasing number of alternative ways of conducting interviews using technology, for example through email, telephone, video conferencing or instant messaging. According to Flick (2014), online interviewing can be arranged in a *synchronous form*, that is, getting in touch with your participants where you can directly ask questions and receive answers while being both online. This kind of exchange is the closest to face-to-face interviews. Another way of sharing questions and answers is *asynchronous form*, which is mostly done in the form of e-mail exchanges. The decision of employing synchronous or asynchronous data collection will impact the nature of the qualitative data. An asynchronous form of interview is preferable when the participants have difficulty to schedule a real time conversation and they can best respond at their own free time (James and Busher, 2006; Moloney et al., 2003; O'Connor et al., 2008; Sweet, 2001). On the other hand, collecting synchronous data has the advantage of mimicking a real time conversation (O'Connor et al, 2008) but in a more modern and convenient way without the interruption of the face-to-face interaction. While I considered the use of asynchronous data, I decided to employ face-to-face interviews in Stage 1 and telephone interviews in Stage 3, thus both were synchronous data. Rumsey et al. (2016) similarly employed both face-to-face and telephone interviews to elicit feelings, perceptions and views about the IELTS. This method was viewed as preferable, as the synchronous form of interview would allow me to directly ask questions to participants. Furthermore, in the event that participants would require clarification on any of the questions, I would be able to immediately provide an explanation.

#### **4.9. Interviews: Stage 1 and Stage 3:**

The data collection occurred in three sequential stages, namely: a face-to-face interview (Stage 1), a questionnaire (Stage 2), and a telephone interview (Stage 3).

It was decided to conduct two sets of interviews as a result of the quality of the data obtained from the face-to-face interviews. These initial interviews were loosely structured and responses were relatively brief. However, once the questionnaire data had been analysed, it became evident that additional data would be useful to gain greater insight into teachers' views on the OOPT, the modular system and factors linked to students' achievement.

The telephone interviews generated additional useful data for several reasons. Firstly, they were more structured, which led to longer, more detailed responses. Secondly, questions in the interview schedule were informed by the questionnaire findings. For instance, the questionnaire illustrated the large number of participants who were unfamiliar with the OOPT. In response to this, I decided to provide participants with sample questions taken from the OOPT so that even those who were unfamiliar with the test could at least comment on their view of the sample. Furthermore, as conducting telephone interviews was significantly less expensive than travelling to Saudi Arabia for a second time, it was easier to work around the participants' availability, which increased the number of teachers willing to take part.

As the study aims to explore teachers' beliefs about the OOPT, the ELI modules and students within the ELI, there is a strong reliance on the qualitative data. Interviews can provide the researcher with 'open-ended information', which 'allows the participants to supply answers in their own words' (Creswell and Clark, 2007, p. 6). A further advantage of collecting interview data is that it 'provides one-to-one interaction between you and the individuals you are studying' (Tashakkori and Teddlie, 1998, p. 102).

Interviews are regarded as very effective tools in qualitative data as they can help the researcher understand 'other persons' (for example, the views and beliefs held by others) and 'constructions of reality' (such as the way phenomena are perceived and constructed by individuals) (Jones, 1985, p. 46). Unlike other approaches, 'when done well' interviews are

‘able to achieve a level of depth and complexity that is not available to other, particularly survey-based, approaches’ (Byrne, 2004, p. 182). According to Mason (2002), researchers can use interviews ‘to help them make sensible, intellectually compelling and systematic interpretations and judgments’ (p. 69).

As this research seeks to explore the complexity behind teachers’ views on the OOPT, the ELI modules and its students within this unique educational context, it was necessary to employ a data collection method, which would be able to achieve the required level of depth to grasp the complexities of the phenomenon. Interviews with the teachers and lecturers of the ELI thus played a major role in the investigation and helped illustrate some of the recurring issues relating to the OOPT and the ELI modules, as well as factors relating to students’ background and preparedness.

The interviews followed a semi-structured design, which offers a compromise between structured interviews, in which a strict interview protocol is employed, and open interviews, which are guided by a broad topic (Dörnyei, 2007). Semi-structured interviews are guided by a list of questions, but the manner and sequence of the questions remains flexible, and most importantly semi-structured interviews allow both the researcher and the participant to explore certain topics in greater depth (*ibid.*). In semi-structured interviews, the interviewer would ask ‘the same questions of all the participants, although not necessarily in the same order or wording, and would supplement the main questions with various probes’ (Dörnyei, 2007, p. 136).

Both interviews were guided by an interview schedule, however in Stage 1 the questions were broader and not as refined. The stage 3 interviews, on the other hand, drew upon the themes covered in the questionnaire and was much more structured and more directly linked to the research questions. By this stage I had already analysed the data from the Stage 1 interviews and the questionnaire, and thus I had both developed as a researcher and gained a greater understanding of the complexity of the phenomenon.

At the beginning of each interview, I offered participants an overview of the research questions and a skeleton structure of the semi-structured interview (Taylor and Bogdan,

1998). The initial questions were quite general and served to develop a rapport between the participants and the researcher. This is an important step as it allows the participants to feel confident in providing responses. Cohen, Manion, and Morrison (2008) stress the importance of establishing an atmosphere in which participants feel comfortable enough to openly discuss a topic.

While there may have been advantages of conducting the interviews in the teachers' native language, I decided to carry out all interviews and design the questionnaire in English. The rationale for this was supported by a number of factors, including the awareness that Arabic was not the native language of all teachers in the ELI and therefore conducting the interviews in Arabic may have advantaged some while disadvantaged others. A considerable number of teachers in the ELI are non-Saudi and, for some, Arabic is a second language with a minority for whom English is their native language. The choice to conduct the interviews in English was also helpful in ensuring consistency across the research instruments. It should be acknowledged, however, that the choice of language employed by the researcher could affect the richness and trustworthiness of the data. In a publication, which draws from the *Researching Multilingually* project, Holmes et al. (2013) point out that 'the complexities and possibilities of researching multilingually are not extensively covered in research training nor widely discussed in the research methods literature.' (p. 286). One of the issues highlighted here is that 'language choice involves the importance of building trust (p. 292). The example is given of a researcher who had to study a dialect spoken in the south of Italy in order to 'establish trust' and 'gain acceptance into the community' (ibid.). With reference to this particular study, because I was already an insider to the community, most of the participants were already familiar with me. There was, however, a scope for the use of Arabic in order to build trust. Arabic was used when contacting participants via email to inform them of my study and invite them to participate and also before the interviews officially began. This allowed participants an opportunity to for ask clarification in their native language prior to commencing the interview.

Collecting the data in English had some clear advantages, including facilitating the transcription and coding process, as there was no need for translations, as well as avoiding the challenges posed by translation.

The choice of conducting the interviews in English was strongly welcomed by the participants, who saw no reason for the interviews to be conducted in Arabic. Perhaps surprisingly, I felt that this decreased the power distance between myself as a researcher and the participants and improved the rapport with participants as I was acknowledging their competence in English. It should be emphasised that the decision to conduct the interviews in English was made together with the participants. Participants were asked for their preference and they unanimously chose English as their preferred language for the interview. With regards to positionality, as a doctoral student and an established member of staff within the ELI, many of the teachers would have regarded me as a senior member of staff i.e. hierarchically above them and thus in a position of authority.

As previously mentioned, Saudi Arabia is very much a hierarchical society and subordinates are generally not involved in decision-making processes taken by those in leadership roles, however my research gave teachers a voice on assessment. Agreeing on a preferred language for the interviews was important to ensure that participants were regarded, as much as possible, as equals in an attempt to reduce the power distance between myself and the participants. While conducting the interviews in Arabic might have resulted in more sophisticated responses for some participants, it is possible that some participants may have interpreted such a decision as a criticism of their competence in English.

#### **4.9.1. Administrating the interview:**

Interviews in qualitative research differ in methodological features, such as length, style or the number of participants (King, 2004). In this study both face-to-face and telephone interviews were employed, as these data collection methods allowed more flexibility and freedom to pursue certain topics or questions in greater depth.

In the data collection phase, I was able to interact with the lecturers in a relaxed and informal manner, as most of them were colleagues I knew well. One of the advantages of this informal context, was that I was able to ask follow-up questions based on their responses and, where relevant, I modified questions according to their previous response. While the interviews were semi-structured, modifying the questions slightly allowed me to adopt ‘a more personal

approach' to each interviewee (Turner, 2010, p. 755). The lecturers were all interviewed individually, and responses were recorded and transcribed (see appendix 7).

One problem that I encountered with face-to-face interviewing was that some of the interviewees had little to say or gave brief and shallow answers. Corbin and Strauss (2015) related this problem to interviewees being uncomfortable. It is of course possible that some participants just felt that they did not have much to say. In such scenarios, the interviewer should have some backup questions, which can be used to give participants a little nudge to speak (Corbin and Strauss, 2015). This experience prompted me to re-examine the interview questions I employed in Stage 1 and develop a much more structured and detailed interview schedule for Stage 3 (see appendix 3). This effectively increased the length and depth of the responses.

The Stage 1 interviews ranged from five to twenty-five minutes as some teachers provided very brief responses. Stage 3 interviews, on the other hand, lasted from twenty to forty-five minutes, perhaps due to the fact that they were guided by a detailed interview schedule, produced longer and more detailed responses, thus providing the richness and depth characteristic of qualitative data. To reduce brief and shallow answers, King (2004) suggested pausing for few seconds before moving on to the next question, which can give the interviewee the impression that you need to hear more on the subject.

Eleven of the interviews in this study were conducted face-to-face (Stage 1), and took place at KAU at the ELI, where I usually work. The face-to-face interviews were conducted in quiet rooms, free from any distraction. Most of the interviewees were colleagues of mine, whom I knew personally and agreed to be interviewed. To familiarise lecturers with the current study, a summary of the main objectives of the research was introduced and explained before commencing the interviews. All participants were aware of the interview procedures and were given assurance about ethical principles regarding voice recording. The eleven face-to-face interviews were digitally recorded and then transcribed.

As the Stage 3 interviews took place over a year after the first data collection stage, I did not have sufficient time to arrange a second visit in order to carry out further face-to-face



interviews. However, I wanted to ensure that Stage 3 interviewees would be given the freedom to add more details or ask questions during the interviews, in a similar vein to the face-to-face participants in Stage 1. Furthermore, it appeared that some Stage 1 interviewees felt uneasy answering questions face-to-face, thus suggesting that some participants may actually prefer a telephone interview. Chapple (1999), for instance was sceptical about the data she would obtain through telephone interviews, because she had always believed in the importance of face-to-face interviewing, yet she found that her data were unexpectedly rich. Chapple (1999) added that when compared to in-person interviews, telephone interviews provide a number of advantages including decreased cost and decreased space requirement.

Furthermore, telephone interviews allow participants to remain on ‘their own turf’ (McCoyd and Kerson, 2006, p.399), decrease social pressure and increase rapport (ibid.). Carr and Worth, (2001, p. 521) described telephone interviews as a ‘versatile data collection tool’ because of the advantages it provides. In the present study, telephone interviews were viewed as a very effective data collection tool.

#### **4.10. Questionnaire:**

The greatest advantage questionnaires offer is their efficiency as they can collect a large amount of data without requiring much time, effort and financial resources. By administering a questionnaire to a group of people, the researcher can collect an extensive amount of information from a large number of participants in a short time and with little personal involvement (Dörnyei, 2007, p. 115). Questionnaires are widely used also to obtain a numeric description of attitudes, trends, or opinions among a population by investigating a sample of that population (Creswell, 2009).

Questionnaires often include both closed-ended and open-ended questions. Unlike closed-ended questions, ‘open-ended questions include items where the actual question is not followed by response options for the respondent to choose from but rather by some blank space (e. g., dotted lines) for the respondent to fill’ (Dörnyei, 2003, p.47). Because there is little personal involvement on behalf of the researcher, participants might feel more at ease answering questions and not feel compelled to provide answers (Nortier, 2008). Munn and

Drever (1990) pointed out three advantages of using questionnaires, namely being time efficient, assuring anonymity, and ensuring that questions are standardised. Additionally, questionnaires can reach a larger number of potential participants in a relatively short time (Munn and Drever, 1990). The absence of the researcher ensures anonymity, which may increase the truthfulness of the responses (ibid.).

By using questionnaires ‘the researcher can be expected to have a fairly advanced understanding of the issue of topic being investigated. That is certainly the case if one is asking about questions involving rating scales; ‘To what extent do you agree that ...’ which can be answered on a strongly disagree to strongly agree rating’ (Tymms, 2017, p.175). Questionnaires can also be a useful tool for measuring things such as ‘motivation’ or ‘attitudes’. In such cases, ‘questionnaires seek to answer questions about people’s feelings, attitudes and perceptions, having first decided what kind of attitudes and perceptions are relevant and valued’. Where open ended questions are employed, ‘the questionnaire is less constrained’ and it is used ‘to develop a structure or theory through the analysis to the responses but even there, the questions that need to be asked need to be based on prior knowledge’ (ibid.).

There are, however, also disadvantages of employing questionnaires. While little personal involvement may increase the trustworthiness of responses, this also means that the researcher is often not available to answer any questions participants may have. Furthermore, if only closed items are used, the questionnaire may lack coverage or authenticity, while using exclusively open questions may result in very brief responses (Cohen et al. 2011, p. 354). Furthermore, unlike interviews, which can be carried out at an appropriate speed, ‘questionnaires are often filled in hurriedly’ (ibid.). There is also the risk that not enough questionnaires are returned. Where samples are too small, this can make statistical analysis impossible to carry out and therefore affect the extent to which the researcher can use the data (ibid.). Low response rate is often linked to the method employed when administering the questionnaire.

In constrained situations, such as in a classroom setting, it is possible to obtain very high response rates, even as high as 100% (Tymms, 2012), however questionnaires are often not

collected in such a controlled way. They may be sent out by post or made available online. In this study, teachers were asked to complete the questionnaire in their own time, so that may have also affected the response rate. While it is acknowledged that ‘low response rates threaten the validity of the information’ in the social sciences it is not uncommon to see response rates below 60% and at times studies with response rates ‘as low as 20%’ are published (ibid., p. 236). The following section describes the pilot phase of the questionnaire.

#### **4.10.1. Piloting the questionnaire:**

Piloting the questionnaire before administering it to a wider population, constitutes an important phase, as it allows the researcher to test the effectiveness of the instrument and identify any wording issues, which may affect the comprehensibility of the items (Cohen et al. 2011, p. 655). It may also highlight the multiple ways participants may interpret the same item. If this emerges as an issue in the pilot study, then items should be reworded to avoid ambiguity.

Piloting the questionnaire has several functions, in particular it can increase the reliability, validity and practicability of the questionnaire (Oppenheim, 1992; Morrison, 1993; Wilson and McLean, 1994). Bell (1999) suggests all data-gathering should be piloted in order to check that questions and instructions are clear. The pilot stage may also highlight items, which may need to be removed, as they do not yield useful data (p.84). Hence, the pilot study also serves to improve the quality and effectiveness of the instrument.

Dörnyei (2007, p. 112) suggests asking 3-4 trusted colleagues or friends to go through the items and provide helpful feedback and so, the questionnaire was piloted with three colleagues working at the ELI, to ensure any necessary adjustments had been made before distributing the instrument to the wider population.

These colleagues were lecturers at the ELI, all of whom were currently PhD students. They offered to complete the questionnaire at their convenience. I chose these colleagues as potential participants, because they shared similar characteristics with the participants included in the main study. Their suggestions were helpful and minor changes were made to the questionnaire as a result of their feedback, such as removing some of the items and re-

wording others to improve comprehensibility. The results of the pilot study also gave me some insight into how participants understood the questions, which provided me with an indication of what some of the responses may look like.

#### **4.10.2. Designing the questionnaire:**

The questionnaire included both closed ended and open-ended items (see appendix 6). The closed-ended items were Likert-type questions asking participants to what extent they agreed or disagreed with the statements. It included a brief explanation of the objectives of the study with the aim of eliciting teachers' beliefs on the OOPT, the English language modules and students in the ELI. Prior to taking part in the questionnaire, participants were assured that their responses would be handled with confidentiality. The Likert-type scales are commonly used especially when investigating or examining attitudes. The five-point Likert scale, (strongly agree / strongly disagree) was used to examine teachers' beliefs about the OOPT and students in the ELI.

The questionnaire consisted of twenty-eight questions, which can be categorised into two parts: the first part consists mostly of closed-ended questions with a few open-ended ones. These types of questions are classed as 'personal factual questions' because they 'ask the respondent to provide personal information, such as age, occupation...and so on' (Bryman, 2001, p. 146). The second part of the questionnaire used Likert-type scale questions. This is a multiple-item measure that 'measures intensity of feeling about the area in question' (Bryman, 2012, p. 166). The scale had five items, ranging from strongly disagree to strongly agree. According to Pallant (2007), 'this type of response gives you a wider range of possible scores, and increases the statistical analyses that are available to you' (p. 9). The following section discusses the approach adopted towards the data analysis of the interviews and questionnaire.

#### **4.11. Analysing the questionnaire and interview data**

The interview and questionnaire data were analysed in the order in which they were collected: Stage 1 (the face-to-face interviews), Stage 2 (the questionnaire) and Stage 3 (the telephone interviews) The questionnaire data were saved electronically and statistically analysed with SPSS version 24, with the aim of identifying patterns in the data and test for statistical significance between factors.

The answers to the Likert-type questions on the questionnaire were coded according to the scale 1 (strongly disagree), 2 (disagree), 3 (do not know), 4 (agree), 5 (strongly agree). There were also personal factual questions, which included information such as age, occupation and years of experience.

The main statistical procedures employed to analyse the data were descriptive statistics, correlation analysis and a Mann-Whitney test. Descriptive statistics were employed to provide a summary of the questionnaire results. The mean and standard deviation facilitated the analysis of the Likert-type items. For example, the descriptive statistics highlighted the fact that the majority of respondents were unfamiliar with the OOPT. A Mann-Whitney test of comparison was therefore carried out between the two groups of participants, i.e. those familiar with the OOPT and those unfamiliar with the OOPT.

Factor analysis was also considered, however due to the small number of questionnaires returned, it was decided not to carry out the test. Tolmie et al. (2011) emphasize that sample size is the most important issue to consider when conducting factor analysis. Tabachnick and Fidell (2001) argue that 200 cases would represent a fair sample size, and 300 would be regarded as a good number. They stress that with a sample of less than 200 it would be difficult to ensure that ‘the observed correlations are not being substantially influenced by random errors’ (in Tolmie et al., 2011, p.174-175).

Quantitative analysis was particularly useful to illustrate the demographics of the participants, identify the points on which the majority of participants agreed or disagreed and to expose the difference in views between participants who were familiar with the OOPT and those

who were not. For example, the statistical analysis was able to identify differences in opinion on certain items on the questionnaire depending on teachers' familiarity with the placement test. This result, among others, helped inform the qualitative analysis, which further explored the difference between the two groups.

The approach towards the analysis of the interview data was informed by the framework described in Miles and Huberman (1994) and Miles et al. (2014). The framework consists of three components: data reduction or condensation, data display and drawing and verifying conclusions. The process of data reduction takes place continuously throughout the analysis. In the first phase, it happens through coding, summarizing, and finding themes. In the later stages, 'it happens through conceptualizing and explaining, since developing abstract concepts is also a way of reducing data' (Punch, 2005, p. 198).

The process of data condensation aims to sharpen, sort, focus, discard, and organise data in such a way that 'final' conclusions can be drawn and verified (Miles et al. 2014). Qualitative data can be transformed in a number of ways through selection, summary and paraphrase and so on (Miles et al. 2014).

The second component, namely data display, is an essential process in qualitative analysis. Qualitative data is generally bulky, and the display of data helps organise and summarise data through diagrams, charts, tables, or networks. According to Punch (2005), it is 'the basis for further analysis' (p. 198). In the analysis phase I used tables to organise information and categorise data according to key themes, which emerged. This process helped reduce the data and exclude unwanted and repeated items, without losing significant information. Miles and Huberman (1994, p.11) suggest that tables offer 'immediate access' to wanted themes. Looking at displays helps the researcher identify patterns in the data and decide either to analyse further or take action based on that understanding (Miles et al. 2014).

The third component, namely drawing and verifying conclusions refers to the interpretation of the data. Interpreting the data takes place throughout the data collection and analysis phase. At first, the researcher begins by noting patterns, explanations, causal flaws and propositions. The researcher should hold these conclusions lightly, and maintain openness and scepticism,

however initial conclusions are still there, vague at first, then increasingly explicit and grounded (Miles et al. 2014). The ‘final’ conclusions may not appear until the entire data collection process is over. This will depend on the coding, storage, and retrieval methods used; the sophistication of the researcher, and any necessary deadlines to be met. (Miles et al. 2014).

The approach towards the data analysis was further informed by the concept of *thematic analysis* (described in the following section), which was employed in the coding process.

Approaches towards analysing qualitative data are incredibly diverse, complex and nuanced (Holloway and Todres, 2003). Braun and Clarke (2006) argue that thematic analysis should be regarded as a foundational method for qualitative analysis and should be the first one that researchers should learn, as it provides core skills, which can be employed for carrying out a number of other forms of qualitative analysis.

Thematic analysis is defined as ‘a method for identifying, analysing, and reporting patterns (themes) within data’ (Braun and Clarke, 2006, p. 6). Braun and Clarke (2006) propose a 6-phase guide to doing thematic analysis; the six phases are the following: 1. Familiarising yourself with your data, 2. Phase 2: Generating initial codes, Phase 3: Searching for themes, Phase 4: Reviewing themes, Phase 5: Defining and naming themes and Phase 6: Producing the report. This 6-phase guide informed my approach towards the analysis of the interview data. Firstly, I immersed myself in the data in order to familiarise myself with the depth and breadth of the content. I then began to generate initial codes with an understanding that these might change once my analysis would become more refined. In the third phase, I searched for themes, which involved going through the initial codes and re-focusing the analysis at a broader level of themes, rather than codes. This involves sorting the different codes into potential themes and grouping the coded data extracts within the identified themes. In the fourth phase, I reviewed the themes and made changes to some of the previously identified ones. For instance, some of the initial themes did not have enough data to support them. In the fifth phase, I defined and further refined the themes that will be presented in the analysis and further analysed the data within those themes. The final phase involved writing-up of the thematic analysis in a detailed way so that the reader is convinced of the merit and validity of the analysis.

The 6-Phase guide to thematic analysis proposed in Braun and Clarke (2006) therefore guided my approach towards the interview analysis. I have included a sample of my initial coding in the appendix (see appendix 8), the process of searching for themes (see appendix 9), which I carried out on a hard copy of the interview transcripts and finally a sample of the more refined themes (see appendix 10). The final themes are also illustrated in figures in chapter 5.

#### **4.12. Reliability and validity**

Reliability generally refers to the extent to which findings of a study are ‘independent of accidental circumstances’ (Kirk and Miller, 1986, p.20). Hammersley (1992) similarly defines reliability as follows:

[Reliability] refers to the degree of consistency with which instances are assigned to the same category by different observers or by the same observer on different occasions. (p.67).

Piloting the questionnaire is one way to minimise the risk of low validity and reliability in the main study and increase consistency in the responses.

In qualitative research, however, reliability is understood differently. Moisander and Valtonen (2006) emphasise the importance of making the research process ‘transparent’ by describing the research strategy and data analysis methods in a sufficiently detailed manner’ and by ‘making explicit the theoretical stance from which the interpretation takes place’ (in Silverman, 2009, p.282). Thus, according to some writers, the criteria employed to establish reliability in quantitative research is not applicable to qualitative studies, which seek to explore a phenomenon rather than prove causality.

As this study comprised both a questionnaire and two interviews, there are aspects of this study, which may be more easily replicable than others. For instance, the questionnaire could be replicated with a different set of participants, either within the same university or in other universities in Saudi Arabia or similar contexts. As the focus of the questionnaire is on teachers’ views on the OOPT, the modular system and the students in the ELI, there may be aspects of the questionnaire which are less relevant to different learning contexts. It is



therefore unlikely that, if replicated, the questionnaire would yield the same results, however it would be possible to obtain similar ones, in particular if the questionnaire was replicated in a similar context e.g. another female section of a Saudi government university with a similar student intake, which similarly employs the OOPT for placement purposes.

Validity means that the findings and the methods used to carry out the research are based on the research objectives. With regards to validity, there are two major types relevant to investigations in the social sciences: internal and external (Berg, 2007). For some quantitative research, internal validity refers to whether the conclusions drawn from a ‘causal relationship between two or more variables holds water’ (Bryman, 2016, p.41). External validity, on the other hand, refers to the generalisability of the findings ‘beyond the specific research context’ (ibid, p.42). With regards to qualitative research, internal validity is understood differently. In the present study, for instance, the interviews sought to explore teachers’ views on the OOPT. The internal validity, here, concerns the trustworthiness of the responses obtained from the participants.

#### **4.12.1. Trustworthiness**

Trustworthiness is a particularly important concept in qualitative research and it refers to ‘whether the data analyses, reports and interpretations constitute honest and authentic reconstruction of the research and the knowledge that emerged in the social environment’ (Paltridge and Phakiti, 2015, p.192). This can also be viewed as the internal validity of a qualitative investigation. Naturalistic enquiries are judged according to the criterion of trustworthiness (Lincoln and Guba, 1985). Lincoln and Guba (1985) use specific terms to refer to the four criteria of trustworthiness, namely: *credibility* for *truth value*, *transferability* for *applicability*, *dependability* for *consistency* and *confirmability* for *neutrality*. ‘Transferability for applicability’ may also be understood as referring to the external validity of an investigation, while credibility relates to the ‘arguments and processes necessary for having someone trust research results’ (Greenwood and Levin, 2007, p.67).

With reference to the present study, my position as an insider facilitated the process of building trust and gaining trustworthy responses from my participants. Participants generally expressed interest in the study and held quite strong views on the topic. While I was viewed as a colleague by my participants, it is possible that their responses were influenced by social desirability bias. Social desirability bias refers to ‘the tendency to respond to questions in a socially acceptable direction’ (Kenny, 2004, p. 1044). This may have been a greater concern had I interviewed students, as they may have viewed me as a teacher rather than a researcher. While Guba and Lincoln (1989) argue that trustworthiness in qualitative research can never be proven, criteria for trustworthiness can help guide the researcher towards portraying an honest and authentic picture of the findings, which emerged from the research. Furthermore, while a researcher may feel confident about the results, there is always the risk that not all respondents provide entirely truthful answers.

Maykut and Morehouse (1994) argue that a researcher may feel that trustworthiness has been achieved when they are confident enough to act upon the findings of a research. This suggests some degree of generalisability, which in qualitative research refers to the relevance of the findings beyond the context of the study. In other words, the external validity of a qualitative study relates to what can be said about the rest of the world, while the internal validity lies in whether the conclusions drawn are warranted in the findings.

With regards to the validity of insider research, some have argued that the perspective of an insider could impede critical observation (Wolff, 1950), whilst others feel that insider research can produce valuable insights (Lewis, 1973). I took this into consideration when carrying out my data collection. There are however advantages of insider research, for instance, Unluer (2012) argues that the characteristics of professional relationships can outweigh issues concerning the validity of insider research.

#### **4.12.2. Credibility:**

The term *credibility* refers to ‘the methodological procedures and sources used to establish a high level of harmony between the participants’ expressions and the researchers’ interpretation of them’ (Given, 2008, p.138). Guba and Lincoln (1989) highlight the

importance of ensuring that constructed realities in the researcher's interpretation of the data match the constructed realities of the participants. They suggest three credibility measures, namely prolonged engagement, member checking and peer debriefing. While it was not possible to spend a significant amount of time within the research context, as the ELI is my regular workplace, I was already very familiar with the setting. I also knew most of the teachers as colleagues and had, therefore, already established a rapport with the participants. With reference to member checking, I started carrying out a preliminary analysis of the data while I was collecting it. This consisted of reflecting upon the interview responses, listening to the transcripts and taking a look at the notes I had made and highlighting recurrent themes and responses. This procedure helped me reflect upon my understanding of the interview responses.

Furthermore, a naturalistic researcher should seek the help of a peer for debriefing in order to strengthen the reliability of the findings. It is preferable that this person be a professional who is not a member of the target population and someone who is willing to analyse, check and test findings. The findings, which emerged from the preliminary analysis were discussed with two PhD students (peer checking, Dörnyei, 2007) in order to gain some feedback on my approach towards coding the interviews and analysing the quantitative data.

#### **4.13. Ethical considerations:**

Punch (2009) stresses the importance of taking ethical issues into consideration when carrying out research, which requires collecting data from human participants.

Ethical dimensions specifically relevant to interview data further need to take into consideration 'informed consent, guarantees of confidentiality, beneficence and non-maleficence (i.e. that the interview may be to the advantage of the respondent and will not harm him or her' (Cohen et al., 2011, p. 686).

The study was carried out in accordance to the ethical considerations specified in the British Educational Research Association (BERA, 2011). The study is therefore guided by the ethical conditions and parameters of educational research. The association argues that 'educational researchers should operate within an ethic of respect for any persons involved

in the research they are undertaking. Individuals should be treated fairly, sensitively, with dignity, and within an ethic of respect and freedom from prejudice' (BERA, 2011, p. 5).

Ethical approval was obtained from the University of Durham before commencing any data collection in Saudi Arabia. I also requested consent from the head of the ELI at KAU to ensure that the university would be informed and supportive of my research. In order to ensure anonymity and confidentiality of responses, participants were assured that their names or any information that could identify them would not be included in the thesis and the data would be stored securely. All participants were informed of the purpose of the study and how to take part. Interview participants were also asked to sign an informed consent (appendix 2). In the case of the telephone interviews, informed consent forms were sent via email and signed electronically. Participants were also informed very clearly that interviews would be audio-recorded as some lecturers may not have felt comfortable with their voice being recorded because of cultural beliefs.

All interview participants were also informed of the right to withdraw from the study at any time. With regards to data protection, any digital data, e.g. transcription of the interviews or audio-recordings was kept on a password-protected computer to which only I had access. As for the paper-based questionnaires, these would be kept only for the duration of the study and later destroyed. Under the Data Protection Act, all of the data collected had to be used exclusively for the purpose for which it was initially provided. In order to protect both myself as a researcher and the participants, I discussed the venue where the interviews would take place with my supervisors, before beginning the data collection.

Other ethical issues, which could arise, concern the sensitive nature of information participants may provide in interviews. Some of the information revealed to a researcher, particularly in interviews, might be classed as privileged information (Smyth and Holian, 2008), in which case confidentiality is extremely important.

To maintain confidentiality, it was not possible to provide much information about the participants. As previously mentioned, some participants were reluctant to talk about testing because they feared that this could have had negative implications for their employment within the university. It was therefore important to ensure that responses remained

anonymous and that information relating to participants would not lead to them being identified. For this reason, it was decided not to provide profiles with personal characteristics of each participant.

#### **4.14. Limitations of the chosen approach:**

The study is a relatively small-scale investigation, which comprised sixty questionnaires and nineteen interviews and examined the phenomenon in a specific learning context: the ELI at KAU. One of the limitations of the chosen approach concerns the generalisability of the findings. As the study was relatively small scale and only examined the phenomenon in one particular context, I must be careful in my generalisations. The findings of this study, however, may inform future research investigations in similar learning contexts, for instance within other PYPs in Saudi Arabia.

The study therefore maintains an intrinsic value, in particular because the qualitative data seeks to develop an in-depth understanding of a phenomenon and thus it is more concerned with ‘trustworthiness’ (Dörnyei, 2007, Lincoln and Guba, 2000) than generalisations. Tymms (2017, p.165) however suggests that the aim of educational research should be to improve the quality and effectiveness of education. Hence, whether a study relates to a specific instance or to a whole population, or whether it is quantitative, qualitative or mixed methods, all studies can make a contribution to improving education. In order to increase the trustworthiness of the responses I also considered the threat of social desirability bias (Dörnyei, 2007). As previously mentioned, because I interviewed teachers in a similar role to mine, the risk was minimal. The threat of social desirability bias may have been greater if I had chosen to interview students, as they may have viewed me as a lecturer, which in turn could have affected the trustworthiness of the responses. This was taken into consideration when I decided only to interview teachers, although one of the drawbacks of not including a student perspective is that what we learn about students is limited to the perspective of teachers. However, had I chosen to interview students, the data collection would have required more time, in particular had I chosen to carry out interviews. This is due to cultural reasons and, for some students, linked to a very strict interpretation of Islamic regulations concerning the recording of female voices and the possibility that these may be heard by men.

For these reasons, interviewing students is particularly problematic in this context and the process of obtaining an informed consent can take longer than expected, in particular where students need to seek the approval of their families.

Further limitations of the methodology employed in this study have been previously discussed in this chapter, for instance in relation to insider research.

With reference to the decision to adopt a mixed-methods design, while there were questions asked in the interviews that were not included in the questionnaire, several items on the questionnaire also appeared in the interview schedule. The fact that quantitative and qualitative findings were generally in harmony with each other encourages a belief in the trustworthiness of the data. This suggests that the research instruments were adequate for exploring the phenomenon under study.

#### **4.15. Summary:**

The chapter described the approach and research design adopted in the study. It further illustrated the rationale for adopting a mixed-methods approach and described the research instruments and design adopted for the data collection. The chapter further outlined the approach towards the data analysis and ethical considerations taken throughout the data collection and limitations of the chosen approach. The following chapter outlines the results of the analysis of both the qualitative and the quantitative data

## **CHAPTER 5 DATA ANALYSIS AND RESULTS**

### **5.1. Introduction:**

The chapter presents the results of the investigation drawn from the interviews and questionnaire. The results of the Stage 1 (face-to-face interviews) are presented first, as these informed the writing of the questionnaire. This is followed by the questionnaire results, Stage 2, and the Stage 3 (telephone interviews), which followed up on the questionnaire results. The results are discussed in relation to the four research questions guiding the investigation.

**Research Question 1: How do teachers view the OOPT?**

**Research Question 2: How does the OOPT impact on placement and teaching, in the opinions of teachers?**

**Research Question 3: What are the factors that might affect students' achievement in both the OOPT and during the English module, in the opinion of teachers?**

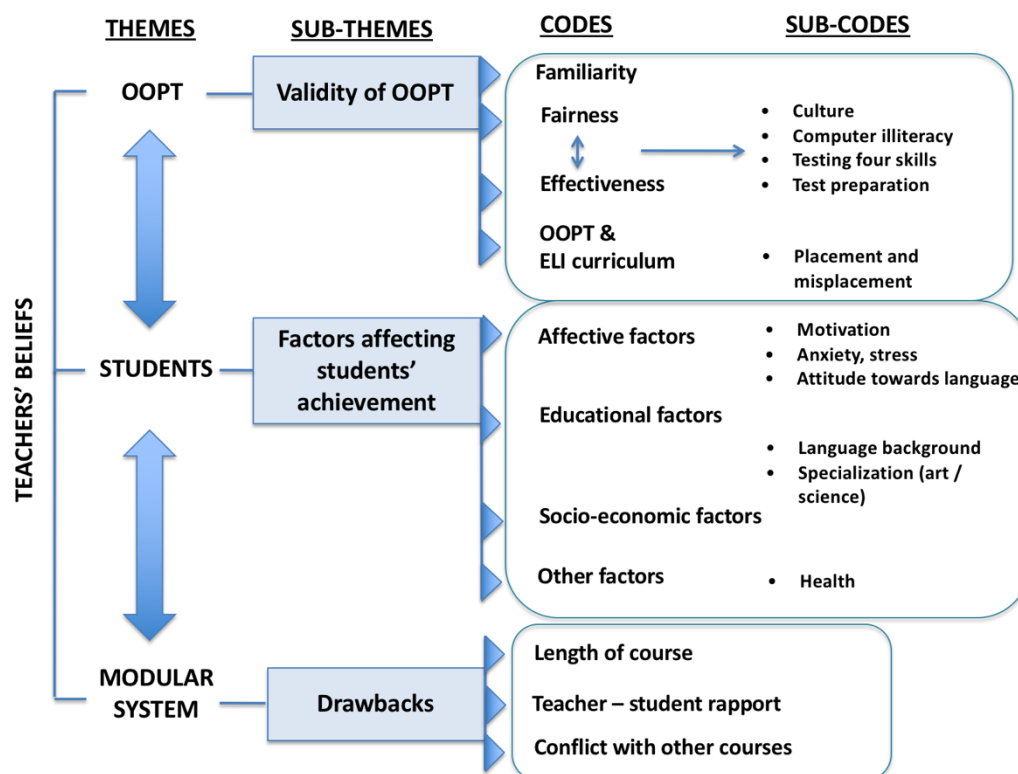
**Research Question 4: How can the system be improved to support students' achievement?**

Results related to Research questions 1 and 2 provide information about the validity of the OOPT, according to Messick's (1989) unified view of validity, while the latter two research questions add to our understanding of the cultural context, the structure of the ELI modules and their relevance to students' language learning during the PYP.

The main objective of the study was to investigate teachers' perceptions of the OOPT in relation to its effectiveness as a placement tool for students at KAU. The investigation furthermore explored teachers' views on the ELI modules and students within the ELI. While teachers' views may contribute to a better understanding of some of the issues related to the OOPT and factors affecting students, which are specifically relevant to the students at this university, it should be noted that the views are specific to this context and are drawn exclusively from the opinions of the participants interviewed.

With regards to teachers' beliefs, firstly sub-codes were identified, which were then condensed to codes and translated into themes and sub-themes. Three main themes resulted from these codes and these are discussed in further detail in this chapter. Figure 10 illustrates the coding process described above.

**Figure 10. Identifying themes**



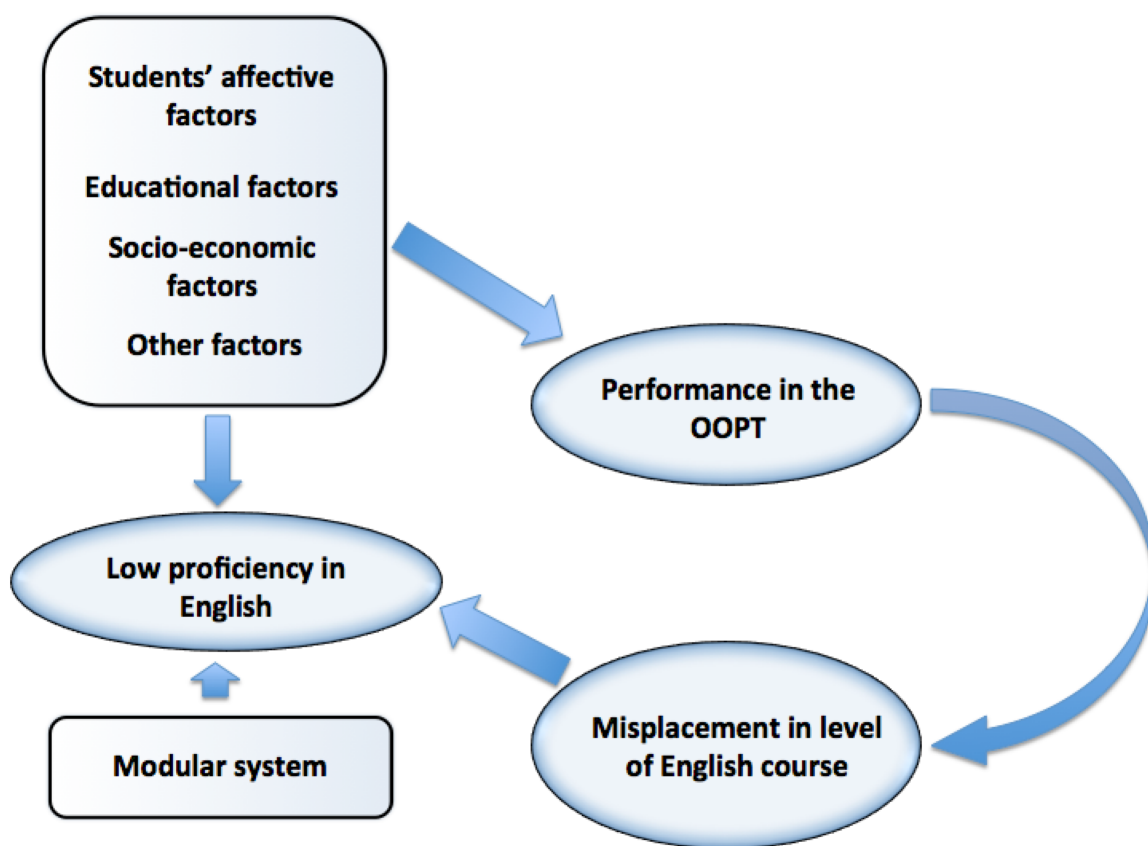
All emerging codes and themes relate to factors affecting students' achievement in the OOPT and in the modular system. All of these are inter-related and must be viewed within the context of the use of placement tests in making assumptions about the level of proficiency that students are likely to attain after an English language course.

The analysis of both the interview and questionnaire data facilitated the identification of key factors that appeared to affect student' achievement both in the OOPT and later in the ELI modules. Teachers made reference to specific factors, which, in their view, contributed to students' performance in the OOPT. This included affective factors, educational factors, socio-economic factors and other factors such as health-related factors. Teachers felt that



these factors could affect performance in the OOPT, which could lead to misplacement and therefore affect their proficiency in English. These factors similarly emerged as having a direct relationship with proficiency, for instance the type of school students attended was often associated with the students' proficiency level. The modular system also emerged as being directly linked to students' proficiency in English, for instance several teachers felt that students did not make adequate progress in these courses. The figure below illustrates how these factors emerged as being linked to students' proficiency in English

**Figure 11. Perceived factors affecting students' achievement and placement in English courses**



The following sections discuss the data analysis of the initial interviews (Stage 1), which served to gain a better understanding of the phenomenon. As very little was known about teachers' views on the OOPT, the students and the modular system at this stage, the interviews were less structured and thus the responses obtained were rather generic. The

following section discusses the results of the first interviews in relation to the three central themes: 1. The OOPT 2. Students and 3. The modular system.

## **5.2. Stage 1 – Analysis of face-to-face interviews:**

The first data collection phase (Stage 1) comprised of face-to-face interviews. This was followed by the questionnaire (Stage 2), which was informed by the interview responses. Subsequently, a second set of interviews (Stage 3) was conducted over the telephone with different members of staff. The telephone interviews had a more structured interview schedule, with questions designed specifically to prompt more detailed responses from the participants.

The first set of interviews helped build an overall picture of teachers' beliefs on the OOPT, the students as test-takers and the modular system. For example, it became apparent that many teachers were unfamiliar with the OOPT. Those who were familiar with the test held different views. A number of teachers felt that the results of the OOPT, at times, led to misplacement into English courses for some students, which were placed either above or below their level of competence in the language; this also had implications for students' proficiency levels. The interviews also highlighted issues with the modular system, which continued to emerge from the subsequent data collection stages.

The following section presents results from the first data collection stage. All interviews were transcribed and then analysed in order to find patterns, which could then be translated into themes. The following section discusses teachers' beliefs in relation to the OOPT, students and the modular system as derived from the analysis.

### **5.2.1. Teacher's views on the OOPT:**

With regards to how teachers viewed the OOPT, only one of the participants in the (Stage 1) interviews specifically stated that she was not familiar with the OOPT, while some others commented that they had not seen the test. Most teachers based their responses to questions regarding the OOPT on their experience of working with students. Participant 3 was one of

the teachers who had not seen the test and pointed out that her comments will be based on what she observed from the students. She felt that ‘when they first take the placement test and **they are placed in a level based on those results**, so in module one where we see that they seem more or less to be in the right level’. Participant 11 had actually taken the OOPT herself and commented that she did not **feel it was culturally relevant** to the Saudi context.

A key question arose, with regards to the OOPT, concerning whether this type of placement test was **fair** for the university’s particular student intake. A number of issues emerged from the interviews in relation to fairness, which included questions about whether the content was culturally appropriate for Saudi students as well as various challenges experienced by students who took the test, e.g. **lack of IT skills or preparation**. With regards to the **cultural content** of the OOPT, some teachers felt strongly that there should not be any modifications, while others suggested modifications to the test in order to make the content more culturally relevant to Saudi students. Participant 1 commented that, unlike the book, which has a special edition for Arab students designed to be more culturally relevant, the OOPT only has one edition for all students internationally.

Aside from cultural content, another issue identified in the interviews was **computer illiteracy**. Most students had just a basic understanding of IT skills, while some had no previous experience because of their socio-economic background, family culture or school experience. Participant 6 argued that ‘students cannot manage computers, they don’t use them, so probably that would affect their grades when they take their exam’. However, this view was not supported by all participants, in fact Participant 5 argued that using the OOPT was a ‘major development’ and that it’s a test that ‘suits this generation’. She added that ‘students are fond of technology, they are used to it, it’s quite easy for them and it’s the modern way of assessment’.

With regards to the fairness of the test, questions also emerged concerning whether all **four skills** should be tested. At present, the OOPT evaluates exclusively students’ receptive skills (reading and listening) and some teachers argued that productive skills (speaking and writing) should also be assessed. Participants were asked about including speaking and writing tests in the OOPT. Participant 6, in this regard, commented that the OOPT doesn’t reflect students’

ability because ‘it does not test speaking and writing’. Participant 2 also agreed that ‘all four skills should be tested’. Participant 11 stressed the importance of including writing in the OOPT and argued that ‘writing makes a big difference in the student’s level and it makes a difference with the teacher when teaching the writing parts, that are in the book’.

Teachers further made reference to **test preparation**, as a contributing factor to test fairness. Participant 9 pointed out that ‘students are not acquainted with these types of exams, they didn’t have the right practice, they are not prepared’. She added that she felt this was unfair to students because ‘they have to have the right skills’ and the ‘right preparations’ to do well in the exam.

Several issues emerged from the analysis in relation to the OOPT’s effectiveness as a placement tool. Some teachers suggested that the test was not accurately measuring students’ competence in the language, thus leading to **misplacement**, while others felt that the OOPT was very effective. Participant 11, commented that the test places students ‘in the right level, but it doesn’t show the ability of the students, it is not deep, it is so general, but beneficial in only placing students in the right level and not in showing their ability’. Interestingly, Participant 11 identifies the purpose of the OOPT, which is in fact designed as a placement tool. Participant 6, also felt that for the majority of cases the OOPT was **effective** in placing students in the correct levels, although there were some exceptions. Participant 10 mentions a case of a student who was not placed in the correct level as a result of her OOPT score. She said, ‘I faced a 102 student who is better to be placed on 101 level. Her comprehension, writing and speaking skills were not of a 102 student’. Participant 9 also argued that ‘there is something wrong with students in level 1 and 2’, suggesting that there might be some placement issues regarding these lower levels.

For students who are wrongly placed for reasons beyond their control, a re-sit test is arranged. Despite the claim by teachers that the test was accurate for placement purposes, all of the current lower level students were being allowed to re-sit the test as they complained that they were placed in a level too low for their competence. Another teacher described the experience of one student, who excelled at English, yet was placed at the lowest level. The student was

allowed to take the placement test again, and this time was put in a higher level, indicating that there may be other cases like this where the reliability of the test is questionable.

One of the participants, Participant 1, who played a central role in the OOPT exams unit, claimed that each year students complain about their levels. Some of them think they have to be in a higher level and others feel that they should be in a lower one. Accordingly, the ELI decided that teachers should evaluate students in the classroom and can move any student whom they feel should be in different level. The participant indicated that in the previous year only 2% of students who retook the placement test ended up with a different level. This would seem to indicate that in the majority of cases, in spite of students' complaints, the OOPT appears to give valid results. Participant 3 made reference to an issue related to the university's policy on the OOPT and said that students in module 104 (level 4) are part of the problem. The OOPT has three different advanced levels (level 4, 5 and 6), however because the university only offers four different levels, all advanced students are combined in module 104. This results in a wide range of abilities, creating some difficulties for the teachers. She also commented that overall the OOPT, 'seems to do its job in placing them'. Participant 2 similarly argued that 'you will end up having classes with multi-level competencies and always get spiky profiles of language skills in one classroom'. She also mentioned that some students, particularly in level 4, do not need her teaching because their language skills are very good. This could be due to the fact that some of the students placed in level 4 may have had a level 5 or level 6 OOPT score. Guessing the answers on the OOPT, was further identified as an issue related to misplacement. Participant 8 commented as follows: 'I have faced a problem when I taught level 104. One student cannot understand anything because she guessed the placement test. That's why she has been in level 4'. Participant 9 also agreed that guessing the answers affected the results that some students obtained on the OOPT. She commented that 'if the student guessed in either level then at that time the placement test would not do its job in placing them in the right level'.

### 5.2.2. Students:

The foundation year students studying at KAU, taught by the teachers interviewed, were all Saudi females with an age range from seventeen to nineteen years. While some students came from more disadvantaged backgrounds, for example they had attended a public school in a rural area where their exposure to English was limited to a few years of classroom instruction, others had attended private or international schools, where English was the medium of instruction for most subjects. In addition, some students had travelled abroad and therefore had very different experiences. A number of factors related to students' background emerged from the interviews as affecting achievement.

Aside from the issues previously discussed in relation to the perceived validity of the OOPT, other factors emerged from the analysis as affecting students' achievement both in the OOPT and later in the English programme. Students' motivation for studying at university emerged as relevant, as well as their choice of specialisation, which, in some cases, was linked to previous knowledge of the English language. Financial background, schooling and health were also found to have an impact on achievement as well as students' previous experience of learning English, which often influenced their attitude towards studying the language at university.

The Stage 1 interviews also highlighted teachers' views in relation to **Arts versus Science** students. Participant 5 argued that 'Science students work harder' compared to those studying Arts. She said they 'worry about their grades and they do well in the exams'. She also pointed out that students who take their language learning seriously 'find it very simple and very basic and they feel sometimes like they need more they need to practise more English so that when they major and study something in Medicine or English literature for example they feel that they need more'. Participant 3, on the other hand, argued that one of the issues in the English language courses concerns the fact that Art and Science students are mixed in the same class and, therefore, it is not possible to teach the course with a specific purpose in mind. She commented that 'even if the teacher had the time and wanted to do something scientific or art or business, some students will not really benefit. So, I just feel it is kind of a waste of this year, we could be doing a lot more'.

Students' choice of **specialization** at university thus emerged as an important factor. Participant 5 commented that 'science students' work harder, they worry about their grades, they do well actually in the exam'. Participant 6 also agreed on this point and argued that science students 'take their education seriously'. She said, 'I've noticed something about science students: they have very good study skills, they know how to study, arts students are more relaxed, and they care less, they are lazier, but we can't generalise this'. One participant described how teachers prayed to be allocated classes with science students, as these students were generally more disciplined and motivated compared to those specialising in the arts. Another of the participants thought that Science students generally had a better command of English as they put more effort into learning as they know they are going to be taught in English in their major. It was proposed that the Arts and Science students should be taught separately. This would provide the Arts students with a general English programme, which was all they needed for their degree courses, as it would be taught in Arabic.

Some teachers also suggested that a course taught for **specific purposes** might increase interest. Participant 3 suggested that for students majoring in science, a course related to their subject area might be perceived as more relevant. She also highlighted, however, that foundation year students 'don't know yet what major or college faculty will choose or accept them' so unfortunately the university can't offer special purpose courses as it used to. What could be possible would be to offer only two types of courses: one for those majoring in the Arts and one for those majoring in the Sciences, however it would not be possible to have a course taught for a specific major e.g. for medical students. She added that 'if we knew these students were going into medicine, we could tailor the material on the curriculum to be something in that field'. This is an interesting point, as it would seem to indicate that the **content of the syllabus** could perhaps increase students' motivation for the course. Participant 9 also agreed that students would benefit from courses taught for specific purposes and commented that 'honestly I think the four modules will only help students to pass the level' and the course would not help develop the skills required for business or finance.

Overall, a number of teachers felt that science students were more motivated and more academic compared to those majoring in the arts. While this may be based, to an extent, on the experiences they had with the students they taught, it also appears to be a generalization, which may be linked to a cultural view of the two subject areas. It is common practice in Saudi high school to recommend students who do not end up with a high GPA to study the Arts, while students who do better are advised to major in the Sciences. This has implications for the way the two areas of specialisations are perceived both by students and teachers.

Teachers also commented on **students' attitudes** towards taking the preparatory year. Participant 7 suggested that the fact that students are required to undertake a foundation year English course 'puts a kind of pressure on them because there are too many quizzes' and therefore students 'think of it as more than a burden than a learning opportunity'. Furthermore, teachers made reference to **previous experience** of learning English prior to university. Some felt that students who had attended **private schools** were often advantaged. This was associated with a greater focus on English in the private school curriculum. Some teachers also made reference to **socio-economic advantage**, which they felt often contributed to greater opportunity to travel to English-speaking countries.

Teachers also made reference to the fact that a large number of students had attended state schools prior to applying for university and had often been raised in rural areas. This was often linked to the family's **financial position** as well as **socio-cultural beliefs**, which further played a role in students' options with regards to their studies. The interviews suggested that the **SES** of the student's family could have much to do with success in the English language modules. Some students had had the opportunity to **travel to English-speaking** countries or had attended private schools in which the medium of instruction was English. Participant 3 pointed out that some more advantaged students had 'the opportunity to get exposed to the language in its purist form'. She added that students from more advantaged backgrounds will often have studied in an environment where 'the learning will be enhanced' and as a result their 'language competence will be better than students who cannot travel or attend private institutions [...] so it does make a difference, it is a harsh reality, but it makes a difference'. Participant 10, however, argued that it was difficult to generalise the extent to



which SES played a role in achievement in the English language modules and commented as follows:

I cannot generalize this because some students are excellent despite their financial background. It might be a personal ability. They are quick learners and pick up the language easily. They depend on the long classes they take every day, they also make use of all facilities provided to them. But honestly the financial background can be an important factor.

Participant 7 also agreed that, while financial status may have advantaged some students, other factors also influenced students' learning. She commented that, while growing up, she knew girls who had attended government schools and sometimes did a lot better than those who attended private schools and therefore felt that 'attitude towards the language' probably played a more important role.

I think it's about the language attitude itself, it doesn't go back to where we were raised, like you can't teach math like you can teach advance math to some people and they wouldn't get it. So, language is more of a skill and a way of thinking more than a privilege to learn.

Teachers also made reference to other factors, such as students' ability to work hard, which they felt could affect students' performance in the OOPT and in the ELI modules. Participant 9 pointed out that aside from students' prior knowledge of English, their ability to learn languages, and whether they are hard working or not, also played a role in their achievement.

Many of the teachers complained, more generally, about the **level of proficiency** of their students. Although students in the ELI generally had been learning English at school for a number of years, many of them were not able to produce very basic sentences. This raised questions about their **language background** and how they had managed to pass their English exams at school. Participant 3 pointed out that students' success depends on 'what they came in with'. She said that 'those who came with a decent background already do fine', however, students who are weaker upon entry 'start off with a lower level' and struggle in the courses. Participant 11 also felt that students' language background continued to play a role in students' learning. 'I think the language background of each student is reflected throughout the four modules. Students who come with background of the English language, whether they studied abroad, took courses or were in a private school, move on easily from levels 1 to 4'. Participant 10, however, pointed out that while students' language background may be

relevantly linked to achievement in the English language modules, this can also depend on other factors. She argued that ‘it is a matter of different personal abilities, styles, and of course motivation’. Participant 2 also agreed that achievement depends on a number of factors, such as ‘how good the teacher is’ and ‘a student’s motivation’.

Interestingly, when asked about language background, one participant made reference to the advantages of **learning English at an early age**. Participant 2 commented that, ‘if somebody got taught English at an early stage, for example it is different for a child who started learning English at five years old, his or her language competence in English will be way better than somebody who has been taught at twelve or thirteen’. Learning English at an early age emerged from this interview as an important factor affecting students’ achievement. While the interview schedule did not include a question on ‘early age’, the questionnaire did include this. The questionnaire results discussed later in this chapter similarly identify learning English at an early age as important.

### **5.2.3. Modular system:**

The interviews placed an emphasis on a range of views on the modular system, with a majority of teachers describing the ELI modules as not being able to effectively develop students’ proficiency. Some of the issues refer to the **short duration of the courses**, low proficiency levels upon entry and **conflict with other courses**. Some teachers also made reference to the **constant change in teaching staff** across modules which may disadvantage students, as they need to become accustomed to the different modes of delivery and teaching styles.

Upon entry, students often had low competency in English and therefore some teachers felt that it was unfair that they should be expected to complete four English modules over a short period of 24 weeks. It was also pointed out that, surprisingly, the English language programme is the only university course designed on a modular basis, while all other university programmes are based on a semester system. Participant 3 felt that unlike the modular system, the semester system is more flexible it allows for some extra time for ‘when it rains, when things go wrong’. In the modular system, ‘if you lose a day, it has a really big

impact on the whole pacing of the curriculum'. She further pointed out that 'there is not enough time to test the material, you know when you stuff two books is just too much, you can teach it, but I don't think students can digest it'. Participant 3 further argued that while the OOPT appears to be effective in placing students in the right level, she felt that students 'don't really progress as fast as the book progresses' before they move on to a higher level. She also noted that she observed a difference between students who progressed from level 1 to level 2 compared to those who entered in level 2 because of their scores in the OOPT. She said that 'you would get students especially in module 2 coming in better and with much higher levels than the book'. This seems to indicate that as a result of the reduced contact time, teachers are not able to cover all materials in the book, which is set for the level that they teach. Some teachers made reference to the fact that students who are placed in a higher level because of their OOPT scores generally have better competency than those, which progressed to higher levels through the modular system.

Some participants, however, looked at the role teachers can play in supporting students' language learning and felt that more could be done. Some argued that having fewer teachers throughout the course of the year would be preferable, as this would ensure continuity across the modules. However, some viewed having a larger number of teachers as an advantage as this would allow students to experience a range of different teaching approaches and be exposed to both native and non-native teachers of English, thus making the experience more varied. Participant 5 commented that one of the good things of the modular system is that the students will benefit from having several teachers. Instead of having two teachers in one year, students will go through four or maybe six teachers because they also have co-teachers for some modules. This will expose students to teachers from different backgrounds.

Among other issues with the modular system, **conflict with other courses** and exams also emerged as affecting students' learning in the programme. This, at times, meant that students had to take an exam within their subject area, which was scheduled at the same time as their English class. Participant 3 pointed out that 'we are the only department that follow the modular system, it creates a lot of conflict, **other exams do not correspond to ours**'. Participant 10 also argued that she didn't feel the modular system was very successful in her university. She claimed that:

Students in the foundation year study many courses beside the English subject, which make them usually busy and do not have enough time for all subjects. This demotivates them and usually they complain about the long hours of classes before or after which they have other classes. Also, it happens that **students have other subjects' exams during English classes** like the Midterm exams, which force them not to attend the English classes. This can add to the idea that all other subjects in the university are semester based not modular based. All these factors can affect the students negatively.

It would appear that one of the issues that the university would need to take into consideration certainly concerns the timetabling of exams scheduled according to the semester system to ensure they do not conflict with the foundation year English language programme.

### **5.3. Stage 2 - The questionnaire:**

The Likert-type questions in the questionnaire may be grouped into three broad categories: 1. Background data, 2. Teachers' views on the OOPT, 3. Teachers' views on students' prior learning experience and 4. Teachers' views on the ELI modules (in particular the modular system). The initial interviews (Stage 1), were helpful in designing items for the questionnaire and they provided an insight into the views of a small number of teachers. Drawing from this initial data meant that the questionnaire items were very specific and, in some instances, served to explore whether similar responses would be obtained across a much larger sample. Question 9 on the questionnaire, for instance, reads: 9. Science students perform better than Art students. This question drew from the comments elicited by some teachers with regards to the difference in performance and ability between Science and Arts students. Other questions, however, drew primarily from the literature review, such as item 8 on the questionnaire: 8. Students need to be prepared for this kind of test (Online test). The Stage 1 interviews did not ask participants to comment on whether they felt students need to prepare for the OOPT. This question was later included in Stage 3, the telephone interviews. In this case the interviews followed up on a finding from the questionnaire.

Teacher responses were coded and inserted into an excel sheet, and these were then imported into SPSS version 24. The major questions were the Likert-type scale questions (19 questions), which were then followed by open-ended and closed-ended questions.

Open and closed-ended questions asked the participants to state their age, qualification, nationality, length of teaching experience and familiarity with the OOPT. The results are shown in the table below.

**Table 3. Summary of participants' background**

AGE	AGE RANGE	NUMBER OF PARTICIPANTS	PERCENTAGE
	25-30	4	6.8%
	31-36	20	33.9%
	37-42	24	40.7%
	43-48	6	10.2%
	Over 49	5	8.5%
QUALIFICATION	BA / MA / PhD	NUMBER OF PARTICIPANTS	PERCENTAGE
	Bachelor's Degree	15	24.6%
	Master's Degree	40	65.6%
	PhD	6	9.8%
NATIONALITY	SAUDI	26	42.6%
	NON/SAUDI	36	57.4%
YEARS OF EXPERIENCE	EXPERIENCE RANGE	NUMBER OF PARTICIPANTS	PERCENTAGE
	1-5	7	11.3%
	6-11	23	37.1%
	12-17	16	25.8%
	18-23	10	16.1%
	Over 24	6	9.7%
OOPT	FAMILIAR / NOT	NUMBER OF PARTICIPANTS	PERCENTAGE
	FAMILIAR	21	35.5%
	NOT FAMILIAR	40	64.5%

The age range of participants was subdivided into five categories: 1 (25-30), 2 (31-36), 3 (37-42), 4 (43-48) and 5 (over 49). The third age group i.e. (37-42) was the largest group of respondents and contributed to slightly more than 40% of the total number of participants. Furthermore, the second age group accounted to slightly more than a third. In addition, age group one, four and five, each accounted to more or less 10% of the 59 participants.

Concerning teachers' qualifications, most of the participants stated that they held Master's degrees, (65.6%), followed by those holding Bachelor's degrees (24.6%) and by those holding PhDs (9.8%). Questions concerning their nationalities, revealed that the majority of the participants were non-Saudi (57.4%) and Saudi citizens accounted for 42.6%. Concerning years of teaching experience, most participants indicated that they had between 6 and 11 years of experience teaching English, accounting for 37.1% of the 62 participants. This was followed by 25.8% of participants, with ranges from 12 to 17 years of experience, 16.1% who had between 1 and 5 years of experience and 9.7% who had over 24 years of teaching experience. Furthermore, with reference to familiarity with the OOPT, 35.5% of participants indicated that they were familiar with the test, while the majority, 64.5% stated that they were unfamiliar with it.

**[The following section reports the analysis of the Liker-type questions, for which descriptive statistics, correlation tests and the Mann-Whitney test were used.**

### **5.3.1. Likert-type items:**

The 19 Likert-type questions used a 5-point scale, ranging from strongly disagree (1) to strongly agree (5). The mean was calculated, and the participants' overall opinion was determined depending on the values shown in the table below:

**Table 4. Likert scale weighting**

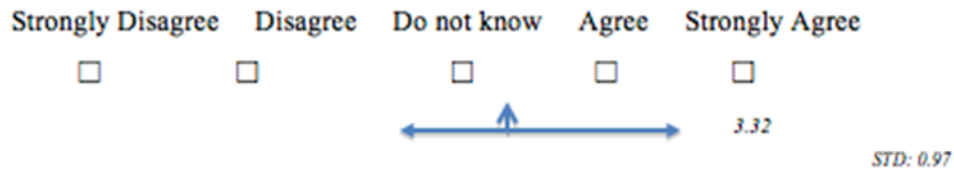
<b>Opinion/attitude</b>	<b>Weight</b>
Strongly Disagree	1
Disagree	2
Do not know	3
Agree	4
Strongly Agree	5

To gain an overall picture of the results, descriptive statistics were carried out with all the responses to the 19 Likert-type scale items. Mean and standard deviation, were calculated to

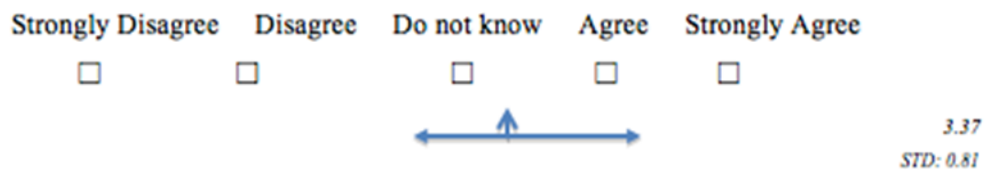
find out the opinion of all respondents. Brown (2014) has identified mean and standard deviation as follows: ‘mean indicates whether the level of agreement to particular questions was generally high or low overall. In contrast, standard deviation indicates how widely dispersed or spread answers were overall’ (p. 151). Figure 12 in the next page presents the mean and standard deviation for all of the 19 Likert-type scale questions. The arrow pointing upwards highlights the central tendency of the responses as indicated by the mean scores, while the horizontal arrow reveals the standard deviation.

**Figure 12. Mean and standard deviation of the 19 Likert-type scale questions**

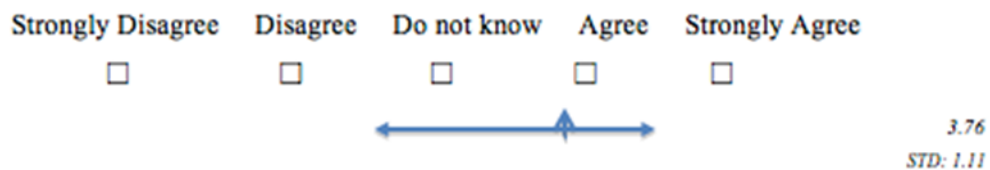
1. I believe that Oxford Online placement test is an effective tool, it places students to their appropriate level.



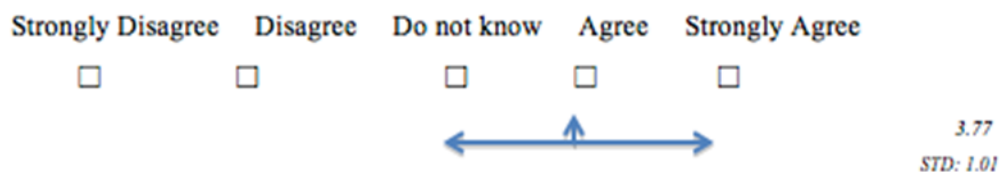
2. I believe that Oxford Online Placement Test reflects student's abilities in the English Language.



3. I think speaking and writing should be included in the Oxford Online Placement Test.

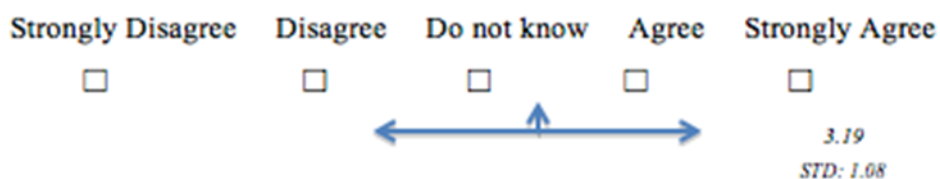


4. I believe that students who were placed in level one struggle in the English module.

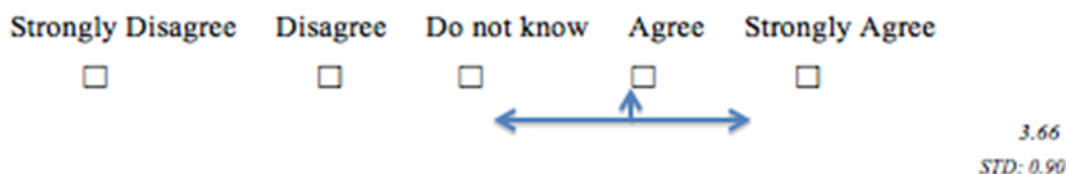




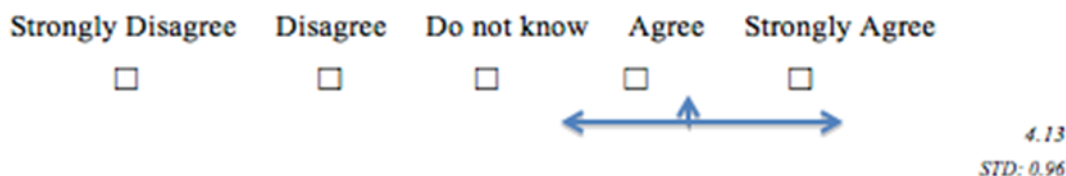
5. I believe that students who were placed in level two struggle in the English module.



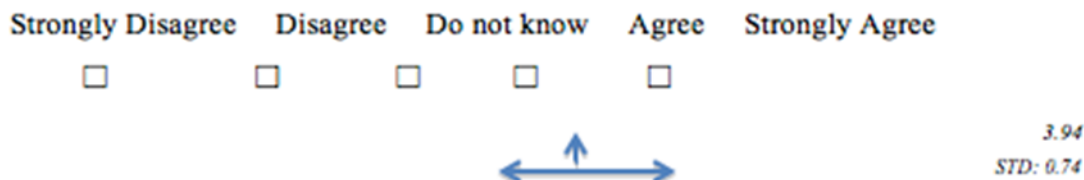
6. I believe that students who were placed in level three or level four succeed in the English module.



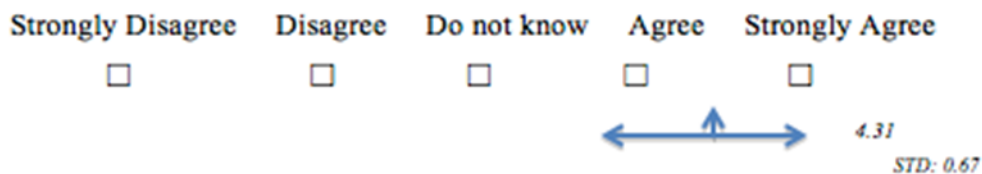
7. Students from private schools perform better in English than those in governmental schools.



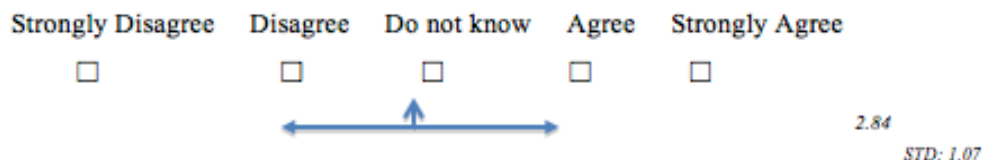
8. Students need to be prepared for this kind of test (Online test).



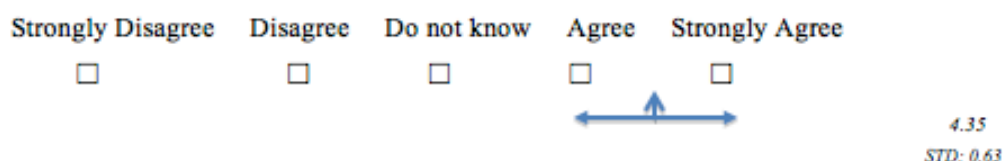
9. Science students perform better than Art students.



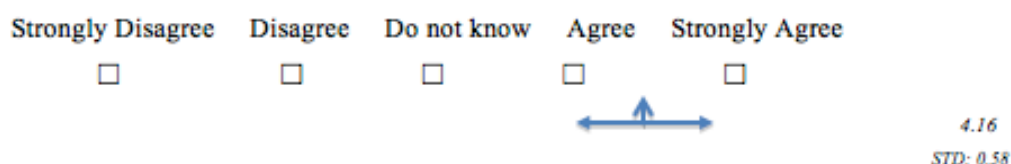
10. Students' financial background reflects their success or failure in the English module.



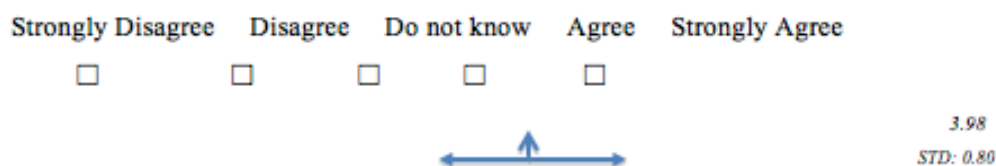
11. Travelling abroad and going to private schools enhance students' language competence.



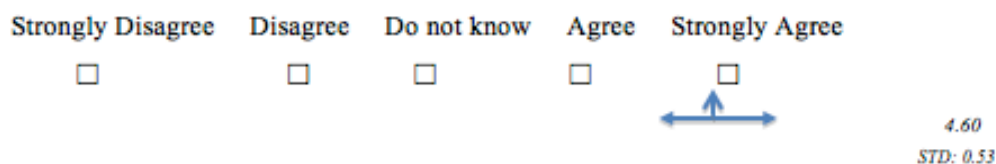
12. Students' success or failure depends on other factors such as personal ability.




13. Language background considered being a major factor in student's success.



14. It is better to learn the language from an early age.




15. Motivation plays a vital role in student's success.

Strongly Disagree	Disagree	Do not know	Agree	Strongly Agree
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
				


4.73  
STD: 0.45

16. The English program will positively facilitate student's future career.

Strongly Disagree	Disagree	Do not know	Agree	Strongly Agree
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
				


4.23  
STD: 0.80

17. I think that the modular system is successful.

Strongly Disagree	Disagree	Do not know	Agree	Strongly Agree
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
				


2.25  
STD: 1.16

18. Having mixed students (art, science) in one classroom are problematic.

Strongly Disagree	Disagree	Do not know	Agree	Strongly Agree
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
				

2.76  
STD: 1.13

19. The Oxford Online Placement test (OOPT) doesn't reflect everything in the student's competence in English proficiency.

Strongly Disagree	Disagree	Do not know	Agree	Strongly Agree
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
				

3.45  
STD: 0.82

**Table 5. Overview of range of responses across the Likert scale**

PARTICIPANTS' VIEWS			
DISAGREE	DO NOT KNOW	AGREE	STRONGLY AGREE
Q17 - modular system	Q1 – OOPT	Q3 – OOPT	Q9 - Art / Science
	Q2 – OOPT	Q4 - misplacement	Q11 - travel + private school
	Q5 – misplacement	Q6 - misplacement	Q14 - early age better
	Q10 - financial	Q7 - private / public	Q15 – motivation
	Q18 - mixed Arts / Science students	Q8 - preparation	Q16 - English programme
		Q12 - other factors	
		Q13 - language background	
		Q19 - OOPT	

As illustrated above, the average response was “DO NOT KNOW” to the following items: Q1, Q2, Q5, Q10 and Q18. On the other hand, participants generally selected “AGREE” in response to the following: Q3, Q4, Q6, Q7, Q8, Q12, 13 and Q19. Furthermore, responses leaned towards “STRONGLY AGREE” for the following items: Q9, Q11, Q14, Q15 and Q16. Participants also largely responded “DISAGREE” to Q17.

Overall, it is apparent from the results that teachers were more familiar with the characteristics of their students, and the factors which can affect their proficiency, than with the placement test itself. Put another way, teachers know their students better than they know the OOPT. Unfortunately, teachers were not sure about the OOPT’s effectiveness or its ability to reflect students’ proficiency or their competence in English.

Nevertheless, they had positive attitudes and agreed on a few points, for example the importance of including **speaking and writing** in the OOPT. They also generally agreed that students in **level one usually struggle** in English whereas those who are in higher levels do better. Moreover, teachers’ answers were on the agreement side with regards to the importance of **preparing students** before the test; they also agreed that students’ **personal ability** could be a major factor that might affect their success in the English modules. Additionally, most teachers were also in agreement with regards to students’ **language background**. They also strongly agreed that **Science** specialists generally performed better compared to those specializing in the Arts. They also strongly agreed with statements regarding factors that might affect students’ language competence and performance, such as, **travelling abroad**, attending a **private school** prior to university, learning the language from

**an early age**, and **motivation**; all these factors might enhance students' language competence and learning. Surprisingly, they strongly agreed with statements that relate to whether the English programme could facilitate students' future careers, while they generally viewed the **modular system** as unsuccessful.

### **5.3.2. Correlations:**

Correlation tests were carried out to further explore the relationships between teachers' responses to the items on the questionnaire and test for statistical significance between the variables. When a correlation shows statistical significance, this indicates that the two variables are interrelated with each other (Dörnyei, 2007). The table below, identifies the correlations between variables, and reports the correlation coefficient (r values) for each of the items. The correlation coefficient, measures the strength and direction of a linear relationship with the value of r measuring between +1 and -1. The closer the values are to +1 or -1, the stronger is the relationship identified in the correlation test.

**Table 6. Correlation test for all variables**

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19
Q1	1	.69**	.043	-.044	-.12	-.19	-.11	.14	-.19	.09	-.23	-.07	-.20	-.19	-.15	.02	.22	.32	-.31
Q2	.69**	1	-.09	-.11	-.26	.67	-.07	.06	-.17	-.17	-.21	-.18	-.21	-.04	-.04	-.03	.22	-.03	.11
Q3	.04	.08	1	-.06	-.20	.02	-.10	.31*	.04	-.33**	-.08	-.03	.12	.21	.12	-.07	.14	-.07	.25*
Q4	-.04	-.11	-.06	1	.47**	-.03	.38**	.05	-.04	.16	.05	.17	-.17	-.01	.03	.04	-.11	-.08	.23
Q5	-.11	-.26*	-.21	.47**	1	-.02	.22	.14	-.05	-.05	.05	.14	-.09	-.07	.04	.11	-.11	.00	.14
Q6	-.19	.07	.02	.03	-.02	1	-.01	-.23	.10	.06	-.07	-.07	.26*	.25	.08	-.10	-.09	.05	.08
Q7	-.11	-.07	-.01	.38**	.22	-.01	1	.17	.16	.25	.30*	.05	.01	-.04	-.00	-.12	-.05	.08	-.05
Q8	.14	.06	.31*	.05	.14	-.23	.17	1	-.11	-.07	.18	.06	-.14	.16	.20	.39**	.23	-.16	-.07
Q9	-.19	-.17	.04	-.04	-.05	.01	.16	-.11	1	.22	-.11	-.19	-.18	.11	.17	-.21	-.39**	-.11	.01
Q10	.09	.17	-.33**	.16	-.05	.05	.25	-.07	.22	1	.40**	-.01	.07	-.16	.02	-.05	-.08	.31*	.04
Q11	-.23	-.21	-.08	.06	.05	-.07	.30*	.18	.11	.40**	1	.01	.36**	.18	.23	.05	-.08	.02	.16
Q12	-.07	-.18	-.03	.17	.14	-.07	.05	.06	-.19	-.1	.01	1	-.14	-.03	.33**	-.09	.02	.06	-.01
Q13	-.21	-.21	.12	-.18	-.09	.27*	.01	-.14	-.18	.07	.36**	.14	1	.17	.11	-.07	.01	-.14	-.03
Q14	-.19	-.04	.21	-.01	-.07	-.25	-.04	-.16	.11	-.16	.18	-.03	.17	1	.56**	.17	-.04	.03	.27*
Q15	-.15	-.04	.12	.03	.04	.08	-.00	.21	.17	.03	.23	.33**	.11	.56**	1	.27	-.04	-.03	.05
Q16	-.02	.03	-.07	.05	.11	-.10	-.12	.39**	-.21	-.05	.05	-.19	-.07	.17	.21	1	.19	-.20	-.21
Q17	.22	.22	.14	-.11	-.11	.09	-.05	.23	-.39**	-.08	-.08	.02	.01	-.04	-.04	.19	1	.13	-.22
Q18	.03	-.04	-.08	-.08	.00	.05	.08	-.16	.11	.31*	.02	.06	-.14	.03	-.03	-.20	.13	1	-.21
Q19	.03	.12	.25*	.23	-.14	.08	-.05	-.07	.01	.04	.16	-.01	-.03	.27*	.05	-.21	-.22	-.21	1

N= 61 for all correlations.

\* p<.05 \*\* p<.01

The table above shows that there was a strong correlation between Q1 and Q2 ( $r=.69^{**}$ ) indicating that the attitude towards the ‘OOPT being an effective tool placing students to their appropriate level’ was in the same direction as the attitude towards the ‘OOPT reflecting students’ abilities in the English Language’. This was the strongest relationship identified by the correlation test and it indicates that those who agreed that the OOPT was an **effective placement** tool also agreed that it **reflects students’ language ability**. A significant relationship was also found between Q14 and Q15 ( $r=.56^{**}$ ) indicating that ‘learning a language from an **early age**’ correlated with ‘**motivation** plays a vital role in students’ success’. A further significant relationship was found between Q4 and Q5 ( $r=.47^{**}$ ), demonstrating that teachers felt that students **struggle** both in **level 1** and in level 2. Items related to students’ **socio-economic** background were also found to be correlated. A positive relationship was identified between Q10 ‘student’s financial background reflects their **success or failure** in the English module’ and Q11 ‘**travelling abroad** and going to **private schools** enhance students’ language competence’ ( $r=.40^{**}$ ) highlighting that the two factors were interrelated.

Interestingly, a positive relationship was also found between Q8 and Q16 ( $r=.39^{**}$ ); this indicated that responses to the item ‘students need to be prepared for this kind of test (Online test)’ were in the same direction as those for ‘the English program will positively facilitate student’s future career’. Furthermore, a significant negative correlation was found between Q9 and Q17 ( $r=-.39^{**}$ ), and hence the attitude towards ‘science students performing better than Art students’ was different from attitude towards ‘considering the modular system being successful’. A moderate relationship was also found between Q11 and Q13 ( $r=.36^{**}$ ) and hence ‘travelling abroad and going to private schools enhance students’ language competence’ was found to correlate with ‘language background considered being a major factor in student’s success’. Furthermore, Q12 was found to positively correlate with Q15 ( $r=.33^{**}$ ) thus highlighting a relationship between ‘students’ success or failure depends on other factors such as personal ability’ and the role of motivation for students’ success. While the correlation test highlighted other moderately significant relationships between variables, when looking at these closely the relationship was not found to be logical and hence these are not discussed here

### 5.3.3. Comparison between familiarity and unfamiliarity groups

One of the questions in the questionnaire survey, asked participants whether they were familiar with OOPT or not. There were only 21 (35.5%) teachers who were familiar with OOPT; whereas 40 (64.5%) teachers were unfamiliar with the OOPT. As the Stage 1 interview and the questionnaire highlighted differences in the opinions of teachers who were familiar with the test compared to those who were not, a comparison test was carried out.

Comparison between familiarity and unfamiliarity groups with OOPT was interesting as it allowed the researcher to explore whether both groups had the same attitude towards each statement or not. In statistics, there are two approaches for analysing data: parametric and non-parametric methods. To use the parametric approach, the data needs to be normally distributed. For the purpose of this study, the data was measured on Likert scale, which is not normally distributed. Hence, the Mann-Whitney test, which is a non-parametric approach, can be used to discover differences between groups. The test result is said to be significant if the p value is  $<.05$ , otherwise it is not significant. Also, for each group, the statements were ranked with respect to attitude, from highest (strongly agree), to lowest, (strongly disagree), using mean, median and SD (standard deviation).

Table 7 illustrates the comparison result between familiarity and unfamiliarity with the OOPT. The results were ordered with respect to the attitudes of familiarity with OOPT. As can be observed in the table below, there was no statistical significance between the two groups except on **two items**:



**Table 7. Comparison between familiarity and unfamiliarity using descriptive statistics and Mann-Whitney test**

Question	Familiarity with OOPT										P-value(Mann-Whitney)
	Yes					No					
	Median	Mean	SD	Rank	Attitude	Median	Mean	SD	Rank	Attitude	
Q1	4.00	3.59	.908	14	Agree	3.00	3.18	.984	15	Neutral	.049*
Q2	4.00	3.82	.501	12	Agree	3.00	3.12	.853	16	Neutral	.001*
Q3	4.50	3.95	1.214	9	Strongly agree	4.00	3.65	1.051	11	Agree	.228
Q4	4.00	3.91	1.019	11	Agree	4.00	3.70	1.018	10	Agree	.370
Q5	2.50	3.00	1.069	16	Disagree	3.00	3.30	1.091	14	Neutral	.297
Q6	4.00	3.82	.958	13	Agree	4.00	3.58	.874	12	Agree	.186
Q7	4.00	4.14	1.014	6	Agree	4.00	4.13	.939	7	Agree	.787
Q8	4.00	3.91	.610	10	Agree	4.00	3.95	.815	8	Agree	.666
Q9	4.00	4.36	.492	3	Agree	4.00	4.27	.751	4	Agree	.933
Q10	3.00	2.82	1.053	18	Neutral	3.00	2.85	1.099	17	Neutral	.915
Q11	4.00	4.36	.492	4	Agree	4.00	4.35	.700	3	Agree	.776
Q12	4.00	4.05	.575	8	Agree	4.00	4.23	.577	6	Agree	.178
Q13	4.00	4.09	.684	7	Agree	4.00	3.92	.859	9	Agree	.521
Q14	5.00	4.68	.477	1	Strongly agree	5.00	4.55	.552	2	Strongly agree	.381
Q15	5.00	4.68	.477	2	Strongly agree	5.00	4.75	.439	1	Strongly agree	.568
Q16	4.00	4.14	.774	5	Agree	4.00	4.27	.816	5	Agree	.425
Q17	2.00	2.43	1.287	19	Disagree	2.00	2.15	1.099	19	Disagree	.503
Q18	2.00	2.82	1.140	17	Disagree	2.00	2.73	1.132	18	Disagree	.712
Q19	4.00	3.55	1.057	15	Agree	3.00	3.40	.672	13	Neutral	.431

The Mann-Whitney test identified a significant difference between responses of group 1 (the familiarity group) and group 2 (the unfamiliarity group) with respect to questions 1 and 2. These concerned the OOPT's effectiveness as a placement tool and whether the OOPT could reflect students' true linguistic ability.

- The familiarity group agreed with (Q1) 'Oxford Online placement test being an effective tool, it places students to their appropriate level', whilst the unfamiliarity

group was neutral. Statistically, this difference in the attitude was significant at  $p < .049^*$ .

- The familiarity group agreed with (Q2) ‘Oxford Online Placement Test reflecting student’s abilities in the English Language’, whilst the unfamiliarity group was neutral. Statistically, this difference in the attitude was significant at  $p < .001^{**}$ .

This was an interesting finding, as it confirms previous results drawn from descriptive statistics, which seemed to indicate a difference in attitude towards the test between participants who were familiar and unfamiliar with it.

#### **5.3.4. Conclusions drawn from the quantitative analysis**

The results of the quantitative analysis illustrate the views of the majority of the participants, observable from the descriptive statistics, as well as the relationships between particular variables calculated in the correlation tests. Furthermore, as a large number of participants were not familiar with the OOPT, a Mann-Whitney test was carried out to compare results for the two groups (those who were familiar and those unfamiliar with the test). While the results of the Mann-Whitney did not identify significant differences in the responses of teachers to most questions, statistical significance was found for Q1 and Q2, which concerned the OOPT, thus highlighting a difference in the responses between the familiar and unfamiliar groups on these two items.

According to the descriptive statistics, teachers tended to agree that speaking and writing should be included in the OOPT. They generally agreed that students in the lower levels struggle more than those in higher ones. Teachers also generally agreed that test preparation was important and that students’ background (socio-economic, the type of school attended and their language background) could all affect success in the English modules. Additionally, motivation and learning a language from an early age were also identified as factors contributing to students’ achievement. Teachers also, surprisingly, agreed that the English programme could facilitate students’ future careers, although they generally viewed the modular system as not helpful. On items 1 and 2 the descriptive statistics do not give a clear

picture of teachers' views on the OOPT as the mean values are calculated using all participants' responses. As discussed previously, the majority of teachers were unfamiliar with the test and hence this affects the mean values of the answers, which highlight 'don't know' as the most frequent response. On the other hand, the correlation test identified a statistically significant relationship between Q1 and Q2, thus indicating that participants who agreed that the OOPT was an effective tool also agreed that it reflected students' language ability. As the correlation test emphasised the relationship between Q1 and Q2 as the most significant, a comparison test was carried out to explore if any difference between responses obtained from the familiarity and unfamiliarity group could be observed. In the Mann-Whitney test, a statistical significance was observed only for questions 1 and 2. This finding confirms that teachers who were familiar with the test, primarily differed in their responses only on the first two questions, which are strongly associated with the content of the OOPT.

#### **5.4. Stage 3 – The telephone interviews**

The second set of interviews was more structured and employed an interview schedule (see appendix 3), which drew upon the findings of the questionnaire and the initial interview. The analysis of Stage 3, resulted in more refined sub-themes, codes and sub-codes (see figure 10). The initial broad themes, which emerged from the analysis of the first interviews, were subdivided into the following sub-themes: Validity of the OOPT, factors affecting students' achievement and drawbacks. The presentation of the data is guided by the identification of these sub-themes and the codes and sub-codes identified in the analysis.

##### **5.4.1. Teachers' perceptions of the validity of the OOPT**

The first theme explored, is the validity of the OOPT, as perceived by teachers. The characteristics, which emerged from the analysis were: familiarity, fairness, effectiveness and the OOPT and the ELI curriculum. All of the characteristics identified were linked, one way or another, to teachers' familiarity with the test and therefore, familiarity emerged as a key issue both in the interview and questionnaire data. From the interviews, it was apparent that the majority of the participants were unfamiliar with the OOPT, which confirmed the questionnaire findings.

#### 5.4.1.1. Familiarity with the OOPT

With regards to familiarity, a large number of participants had not seen the test and hence relied on issues they observed in the levels they taught, which they attributed in part to the OOPT. Some of the participants also refused to comment on the question related to the OOPT because they had not seen the test.

Participant 8 stated that ‘unfortunately as a teacher we do not have any background or any details of the test’. Participant 2, on the other hand, was slightly familiar with the OOPT but commented that she had simply had a look at the screen whilst invigilating; she had no knowledge of the content. Participant 8 commented that ‘unfortunately as a teacher we don’t have any background or any details of the test’. Teachers’ lack of familiarity with tests more generally has also been previously identified in the literature as an issue. In the Saudi Arabian education system, it is normal for teachers to carry out their own assessments of students (Al-Sadan, 2000) and the lack of familiarity with an externally-assessed test is not unusual. However, some participants *did* have some experience of test working with tests. Participant 2 commented that she enjoyed constructing tests, though Participant 6 spoke of her bad experiences with tests when she was a student. Others also said they did not like tests and exams and preferred continuous assessment. Due to their lack of familiarity with the OOPT, participants were shown sample question from the test (see appendix 4), prior to the interviews, to evaluate in terms of its suitability for their students.

Teachers who were familiar with the test, or felt they could comment on it, made reference to its effectiveness while also voicing criticisms. Some teachers felt that one of the issues with the OOPT was the presentation of the test; it was felt that it was not entirely suited to beginners mostly (level 1) as they were required to read instructions and their reading skills might not be good enough for comprehension of what needed to be done. As Participant 1 explained, ‘the beginners have trouble with very basic instructions, so if the instructions are not clear enough, they will not be able to perform the task. It will be nearly impossible if they don’t understand the question’.

The beginners are the ones most likely to have difficulties with the test, as their level of English is simply not high enough. Pictures are helpful in this case, but the questions still need to be clear (Participant 6). This appears to indicate that lack of familiarity with

instructions in English may also have contributed to misplacement for beginner students, as they may not have fully understood what the test was asking them to do.

#### **5.4.1.2. Fairness**

Participants were asked whether they felt the test was fair for all students, or whether there should be any changes for Saudi Arabian students. Various issues emerged from the interviews in relation to fairness in particular culture, computer illiteracy, testing four skills and test preparation.

Most teachers agreed that the test appears to reflect the language abilities of the students for the first module, as those who have been entered for the exam generally manage to achieve. However, problems arise once they go to the next modules as there are many failures in the institute's exams. It is felt that this may not be the fault of the actual OOPT, but is a result of students being pushed too quickly through the modules. This was seen as being unfair to the students, who may believe they have reached a certain level of proficiency, when in reality they are much lower, but it appears there may be cases where teachers pass students when they do not deserve to pass. For students who want to major in English Literature, this could have a significant impact, as it would not have enhanced their level of competence in the language. However, the students themselves may be of the opinion that they are better than they really are, and this may have implications. If they are not able to cope with the specialist degree course, because their English is not up to the right level, then they may lose confidence in themselves.

Participant 1 felt that there should not be any modifications to the OOPT, however emphasised the importance of cultural sensitivity and commented as follows:

This is an English proficiency test. It should be international, it should be generalized, it should be the same anywhere in the world. I shouldn't test students, for example, in Japan differently than from Saudi Arabia and give them a different test; it is a proficiency test, it should be non-biased. It shouldn't be designed for a certain community. Certain things could be taken into consideration, like cultural taboos like homosexual; you don't want to put that in a test in Saudi Arabia, or mentioning alcohol, you don't want to mention things that not ...they should be culturally sensitive but not when it come to the language testing itself. It should be the same everywhere. (Participant 1)

**Cultural content** was mentioned by several participants as being a barrier to understanding the test and some teachers argued that, as a result of its ethnocentric content, the OOPT was not an effective assessment tool. As Participant 4 reports, complaints about the test mainly concern the fact that teachers feel it does not reflect students' actual skills and abilities. She also believed that the OOPT is not an effective tool as it is 'designed on the ability of students to imitate and copy native English-speaking models'. These models should not be used as a frame of reference for non-native English speakers, and it is not satisfactory to compare English being used in multicultural communications with interactions in a monolingual context. According to Participant 4, the test does not take into account multilingual competency. This, however, overlooks the fact that the OOPT is being used simply as a placement test to ensure that students are being placed in a class with others of a similar level of competence. It is not a true test of language competence as such. However, if teachers have negative attitudes towards such tests, then the students themselves may lose faith in the system.

Participant 2 and Participant 5 felt strongly that there should be a special edition for Saudi students, as the OOPT relied on foreign metaphors that no Saudi student would understand. She gave examples such as skiing or rugby, as these were not part of Saudi culture; in particular female students were even less likely to be familiar with this content since most Saudi girls do not practise any sport. The OOPT was not quite right for Saudi students, argued Participant 6, as nowadays they were not so influenced by American or British culture; rather Saudi Arabia should create its own placement tests reflecting its own traditions. Participant 4 similarly held strong views about language and identity and was opposed to testing to assess competence. She also did not believe in the globalization of English as a standard communication language.

I believe we have to use English but with a Saudi Muslim identity. I can use my own language while I am using English; this does not mean I don't speak well, because I do, but I use my own variety with my own accent as I want people to know I am Saudi. I don't want to be mixed up with native speakers from other nationalities. (P.4)

As Saudi Arabia wants to place itself on the international stage, replacing an internationally accredited test may also have an impact on the quality of education that Saudi Arabia would like to achieve. Part of the Saudi government's strategy has been to encourage foreign education providers to set up in the country, in order to deliver educational programmes that

meet international standards (Phan, 2017). Participant 5 further commented on the cultural concepts included in the OOPT test and felt that some of these would not be understood and therefore could have an impact on test results. Participant 4 also felt very strongly that the OOPT was biased against Saudi students, and therefore was not an effective tool. She believed it measured proficiency based on conformity to native English-speaking norms, and students' ability to copy native speaker models; this was not the way to test intercultural interactions.

**Computer literacy** further emerged as an important area, as some teachers felt that students were not accustomed to online testing. Some argued that this could disadvantage students who have weak computer literacy and possibly lead to misplacement because of low marks obtained in the OOPT. Participant 5 argued that one of the main issues was the lack of student computer skills. She spoke of how some did not know how to log in or input the username and password and how teachers had to help them with this.

We do face students who don't even have a computer at home, these students come from the country side. One of the students told me my parents refuse to install any program in the computer, not even word program. Another student told me that her parents refuse to participate in the English forum. So, we do face these issues we always have technician in the rooms, we show and help the student with any technical problems. (Participant 5)

Participant 1 similarly made reference to a student who had no idea how to use a computer and described how shocked she was to find such lack of technical skills with young people nowadays. On the other hand, Participant 2 said that she had never come across any student who had technical issues, as they all knew how to click, use a keyboard and the mouse. Participant 8 also agreed that all of the students knew how to use a computer, but she identified the problem with technical issues occurring during the test; she mentioned that there were not nearly enough technicians available to help out when needed. Young people today are computer-literate, argued Participant 3, as they all had smart phones and used smart devices all the time. She suggested the only problem with the online test was internet-connectivity. It would appear that participants differed significantly in their perspectives on students' familiarity with computers and IT skills in general. This may also be linked to their own personal experiences with the students they taught and issues, which were brought to their attention.

Teachers held mixed views with regards to **testing the four skills**. Some felt that for a placement test to be fair, it should test all four skills, while others disagreed. Assessing speaking in a placement test was ruled out by Participant 1 as being too time-consuming. She described placement tests as ‘an elimination process’ to ‘decide who stays and who goes’. She added that ‘you don’t want to waste so much time on each and every student, and speaking assessments take too much time’. In addition, speaking tests require a number of teachers to conduct them. It would not be the same with writing, as it was suggested that teachers could take the tests home with them to mark. This may however present a problem as many teachers already complain about their workload; if they were not willing to conduct speaking tests, then it is unlikely they would want to mark writing assessments at home.

It was also suggested that Saudi students should not be expected to have speaking and writing skills; Participant 2 explained that they were taught receptive skills in the classroom, and the traditional teaching approaches used in Saudi schools simply did not prepare students other than in ‘grammar rules and reading’. Participant 3 further argued that ‘this would be unfair as it would not reflect their ability or background’. Participant 6 also felt that the OOPT should test writing and speaking skills; she recounted how one of her more advanced Level 4 student was unable to produce a satisfactory piece of English writing or use the past tense when speaking, and this was not acceptable at an advanced level. Participant 5 similarly was an advocate of including a writing section in the OOPT as she said that students in higher levels might know their grammar but do not know how to write a paragraph, however she agreed that assessing speaking might be problematic and not of central importance since it is not a skill that is used outside the classroom.

Speaking is a problem in all levels I don’t think even if we evaluate them we can do nothing about them like we are in this country they are not going abroad we are not using the language outside the classroom we do not communicate or use it all the time until she finishes her first year at university, when she moves to second year if her major needs English she will use it and practise it.

From the view expressed above, it would appear that the student intake, i.e. the students’ socio-economic or cultural and religious background, may also play a role in their willingness or ability to use English in a variety of different contexts outside of the classroom.



Most teachers agreed that speaking should not form part of the placement test, as this was the skill in which most students felt the least confident. While reading lots of books in simple English and listening to English can have a positive effect on students' competence, according to Participant 3, some of the students may actually have a good level of English language but they do not use it, as they do not have enough courage to speak. They are wary of making mistakes and do not want to speak unless they can use the language perfectly. Even though teachers may tell them they learn from their mistakes, they still refrain from speaking in front of their peers. Most of the time the students are embarrassed by their lack of language skills and Participant 5 links this to the cultural context. People tend to make fun of someone who makes a mistake; this can happen even if someone makes a mistake speaking Arabic. According to P. 5, this is what makes people hold back from speaking in class and affects their performance in the language. This reluctance to speak and fear of making mistakes carries over into group work, and even to English studies in the UK or USA. Saudi students will find other Arab students to be in the same group and will speak in Arabic with them, even though they may produce written work in English (P.5). This defies the purpose of grouping students for speaking activities, as they are still discouraged from speaking English. It is argued that students not only enter university with a hesitation to speak English but also continue to be hesitant to use the language with their peers and teachers in their English language courses.

Hence, while some teachers argued that all four skills should be tested in the OOPT, this would not necessarily ensure that students are accurately placed as there will always be learners who have strengths in different skills i.e. some may be more fluent speakers, while others may have a more solid knowledge of grammar.

With regards to **test preparation**, most teachers agreed that students would benefit from preparing for this kind of test. Some teachers commented on aspects of the OOPT, which could lead to misinterpretation, in particular if students had no previous experience with these types of tests.

Participant 8 found that the pictures in the OOPT were misleading and confusing for students, and she also remarked on the type of questions being used. These questions were not the

direct ones that Saudi students were used to, like the exams they had in high school; this test was asking for responses involving inferences. Both the way the test was administered (online test) as well as the wording of the questions and structure of the test presented challenges for Saudi students, who had no prior experience of this kind of testing in school. While the university offers an orientation workshop, which is mandatory for every student prior to taking the OOPT, participants raised concerns regarding its effectiveness in preparing students, in particular because it was found to be too general. Participant 6 noted that students were not being prepared for such tests at school and she suggested that the fault may lie with awareness of the exam. From attending the orientation, students had the impression that the test would be fairly easy as a result of how the test was presented:

Because what we do in the university, they are providing us with a power point and slides to show how the exam is going to be but teachers, to be honest, they don't take this seriously and for that students think that the exam is so easy and don't need to know about it before. When they go to the exam they find it difficult to access and difficult to get familiar with, so I think it is really important to give a very strong preparation before the exam. P.6

This may indicate that more teacher-training is required for teachers to realise the implications of students not being able to understand the test requirements. There appears to be a relaxed attitude on the part of the teachers, and this is transferred to the students, who do not take the time to research information about the OOPT beforehand. This complacency may have an impact on test performance, although Participant 2 was of the view that any orientation simply served to help students move through the test more quickly. Nevertheless, this is also an important factor in any test; if students run out of time, they cannot finish the test and gain needed marks. Indeed, Participant 6 explained, some students were using social media to share the content of the OOPT after they had taken the test. Instead of teachers guiding and supporting students, they were consequently relying on peer support.

#### 5.4.1.3. Effectiveness

Participant 2 commented that she felt the OOPT was an **effective placement tool** but pointed out that she was unfamiliar with the content of the test; she said 'I haven't seen the placement test, I haven't seen the questions'. Others were also satisfied with the effectiveness: Participant 1 was positive about the test as it represented an international dimension: 'I think any international tool will be an effective tool if it is used and recognised internationally. The most important thing is that to be international and non-biased'. The participant highlights

the fact that the test offers an equal opportunity to all students around the world without favouring one particular group. Participant 3 thought that the OOPT was succeeding in its purpose by placing similar students together:

I think it is effective, because otherwise if we don't have that placement test I mean you will put the wrong person in the wrong level, A placement test just puts the right person in the right place, I think, because as we can see when the students who are in level 1 really have the same competence, the same language background, the majority are very similar and will have the same level. P.3

Some teachers felt that the OOPT was not an effective tool because of its cultural content. This links to some of the issues previously mentioned in relation to culture. In particular, Participant 4 argued as follows:

Not effective, this kind of tests measures the proficiency based on conformity to native English speaker norms, basically American and British varieties. These models should offer an idolization for non-native English speakers. But, they should not act as a frame of reference to capture errors and judge proficiency, legitimacy or intelligibility.

Another issue, which was mentioned in relation to effectiveness, was the mismatch between the OOPT results based on 6 CEFR levels and the university's modular system, which only offers four different levels of proficiency. Participant 7 was of the view that the OOPT was effective but not for their context. She argued that 'the placement test is not suitable for the modular system'. Participant 5 also made reference to issues concerning placing lower level students correctly. 'I believe they have placed beneficially, as I said, 2,3,4 are placed perfectly, the only problem is with students who were placed in the first level'. This might be linked to the previously mentioned issue regarding the mismatch between OOPT and the levels offered in the modular system. A further issue concerned the beginner's level course. The level covers a wide range of abilities; it includes all those who cannot achieve the second level. Teachers felt that, because of the wide range of abilities, it is very difficult to teach this group. Participant 5 reported how her students ranged from those who were at the top end of Level 1 to complete beginners. When they move onto the next level, some leave with an A grade, whilst others will be moved on with a fail. This shows the discrepancy in levels, even when placed at the lowest level. P.5 suggested having a preparation course for complete beginners, as she estimated that Level 1 was not suitable for 20% of the students placed in that course.

Sometimes the students seem to be placed in a level too low for their ability, but it is suspected that this is more the fault of the student rather than the test. Some students fail the test deliberately ‘because they want to end up on an easier level and get higher grades at that lower level’ (Participant 1). This is a ploy, whereby students can be awarded higher grades in exams on the lower level, rather than lower grades on the next level. The reason for doing this is that the grades count towards their GPA, whichever level they are, and this boosts their overall academic grading, especially if they want to specialise in science subjects, they need a high GPA. At present, there is nothing that can be done about students who misplace themselves intentionally as the university allows students to select their own levels.

One of the participants, Participant 3, spoke of mixed reasons for students being misplaced. She was teaching the lowest level but had students who were very good and should not have been put in that class. When she asked why they were not in a higher level, they told her the university had simply placed them in a low level because they had not taken the test. One student was actually exempt from taking the OOPT, as she had passed the IELTS, but was still placed in the lowest level. P.3 also reported that some of her students had guessed the answers and she wondered if they were just treating it all as a game, instead of something serious.

There have been individual cases where the placement tests were not effective in being able to accurately assess students’ competence in English. Participant 6 spoke of a student whose mother is American and whose level of spoken English is advanced. This student was placed in one of the lower levels, despite speaking excellent English, and this was because she was not good at grammar. In the classroom, therefore, she dominated the sessions by answering all the questions and not giving the other students a chance to respond or participate. By focusing on the grammar aspect, the OOPT does not take into consideration the varying levels of skills, known as a ‘spiky profile’, and thus it cannot ensure that students will be placed in a level appropriate to their needs.

#### **5.4.1.4. The OOPT and the ELI curriculum**

The interviews also highlighted a perceived discrepancy between the Oxford placement test (OOPT) and Cambridge curriculum (i.e. the textbooks used in the ELI modules). The interview question related to the perceived mismatch between the placement test and the books was only asked in the Stage 3 interviews, because the decision to adopt Cambridge materials instead of Oxford was made in the following academic year and therefore, during Stage 1, teachers were still using Oxford books. Interestingly, while participants from Stage 1 were not asked to comment on this issue, a reference to the curriculum emerged from one of the interviews. Stage 1 Participant 11, when asked about what she thought of the OOPT, commented that she thought ‘the placement was really related to the books, [...] so students who were put by the placement into level 1, the books in level 1 were suitable for them and so on in level 2 and 3 and 4’. It was pointed out that teachers may be receiving students at a specific level in their classes and teaching them from Cambridge materials; these may be written for higher levels than the Oxford materials targeting that level. This was confirmed by Stage 3 Participant 2, who found the two lower levels were compatible with each other, however, the two higher levels in the Cambridge materials were considerably higher than those in the Oxford resources. Nevertheless, Participant 2 felt that the exam should be linked to the curriculum. In similar vein, Bailey (1998) argued that ‘since the purpose of a placement test is to assign students to particular levels of a program, then it makes sense that the content of the test should be related to the curricula of those levels’ (p. 38). The following interview references are all drawn from Stage 3 as there were no further comments on the curriculum from the Stage 1 interviews.

Participant 7 also agreed with the view that the placement test and curriculum ‘should go hand in hand, they should speak to each other they should be designed for each other’. Participant 5 similarly identified the mismatch between the Oxford and Cambridge publishers as an issue; she expressed a very strong view on the subject stating as follows: ‘I think it is a major mistake, students may be placed correctly in the right level, but I am sure if that exam was exactly like the book, the output would be much more specific’. Participant 4, however, disagreed and felt that it was the teacher’s responsibility to design adequate teaching materials and employ the book accurately. ‘If the students were taught the right curriculum and the teacher is teaching her book properly so I think it won’t affect which

exam the student takes. It is not really important to follow the same publisher'. While it may be argued that the role of the teacher is central to an effective learning experience, it may be preferable, both for the students and teachers, to adopt materials from the same publisher in order to ensure consistency.

The following section focuses on factors affecting both students' achievement in the OOPT as well as in the English modular courses.

#### **5.4.2. Perceived factors affecting students' achievement**

**Affective factors**, which emerged from the interviews relevant to the students' learning experience included: **motivation**, **stress** and **anxiety** and their **attitude towards the language**. Affective factors refer to the emotional side of a human being (Hulse, Egeth, Deese, 1981, p.4). Motivation was found to play an important role in students' achievement both in the OOPT and in the English language courses run by the university. Both intrinsic and extrinsic motivational factors emerged from the analysis. Some students valued developing proficiency in English because this was required for their field of study or future career, and hence were extrinsically motivated, while others enjoyed the learning experience and were therefore more intrinsically motivated. Some students may not see the relevance of learning English as they do not see it having a purpose in their future. These students may treat learning English as university requirement and hence not be very interested or motivated to study the language. Other Saudi students, on the other hand, who travel frequently and spend summer holidays in the UK or America, may be more interested in learning English as they know they will have opportunities to use the language. Another motivational factor, which emerged from the interviews, was the choice of specialisation chosen by the student. Those intending to study the Sciences or medicine know they will need English in future.

Participant 5 highlighted the difference in attitudes towards studying English between Science and Arts students. She argued that, unfortunately, for anyone specialising in Arabic, geography or history, English will seem like a waste of time and effort. This can make a significant difference to the way in which students approach learning the language. The Science students are definitely more appreciated by the participants in this study, but it was explained that they needed a higher GPA from school to be able to get into these specialised

courses; those who did not have a high enough GPA were placed into Arts courses, even if they were interested in the Sciences.

There are also a number of students who do not view learning English in a positive light. According to Participant 6, some students are simply not interested in the language. Many of her students argued that their reasons for studying English were just to pass their degree course as they did not like the language. However, P.6 attributed much of this to the fact that students in the PYP had a lot of pressure placed on them and they needed to gain good grades to be allowed to study their specialist subject. Similarly, Participant 7 also argued that students view English as a requirement to be admitted onto their degree programme and are often demotivated to study the language.

Participant 4 identified yet another issue, which may influence intrinsic motivation, namely that of the quality of the learning experience in English courses. She makes reference both to the level of challenge as well as to the need to change the traditional teaching approaches, which focus on the four skills.

Students become more motivated to study and to improve their English when their course seem related to their interests and needs and when the course seem more advanced and challenging and when assessment methods seem more flexible more interactive more dynamic and more fluid, they appreciated when the exam gives qualitative evaluation description for the weaknesses and strengths in each different skill they truly need other than the basic traditional four language skills.

It would appear that current teaching approaches may also need to be re-evaluated in order to identify ways in which language courses can increase students' intrinsic motivation for the subject. This may have implications both for achievement and students' enjoyment of the learning experience.

Aside from motivation, **anxiety and stress** also emerged from the analysis as factors contributing to students' achievement in English courses as well as in the OOPT. Most students experience stress and anxiety throughout their academic years, especially those who are in the foundation year. According to Seipp (1991), anxiety can be a normal reaction to any new experience and, as the university experience is totally different from schooling years' experiences, too much anxiety can affect their learning and proficiency. Participant 7 explained that the atmosphere of the PYP is very stressful, 'students have to handle many

courses and they have to have certain grades in-order to specialise in the major they want. That puts pressure on the student'. The participant suggested that 'if they can take English in a relatively less stressful environment, that would affect their competence and would be positive'. If this stress were taken away, Participant 6 believes they would relax and that would affect their competence; they would then have a more positive attitude towards the language.

Students' **attitude towards language** further emerged as affecting achievement in English courses. Many students lack self-confidence and some are convinced that they will never be able to learn the language as it is too difficult. This is especially true with regards to the speaking component. Participant 5 illustrates how students view developing competence in English like any other subject that they can take home and memorise, but they need to be reminded that learning a language is a skill, not theoretical knowledge. It is unfortunate that some students lose their confidence at school level and they give up hope of learning English before they even start university. In schools, teachers may not have been proficient at English themselves and they have simply encouraged their students to memorise and pass English exams. P.8 suggests that teachers are not sufficiently qualified and this is why they teach in traditional ways; she recommends that they be asked to take professional developments courses to update their skills.

Some teachers also made reference to how student use of Standard English was perceived as an indication of better competency in the language. Participant 4 argued that students who were able to use Standard English were regarded as being at a higher level than others, especially in reading and writing. Furthermore, she suggested that those who used non-standard English were placed at lower levels, even though they may have been 'proficient users of English'. This suggests that some students were able to perform well in reading and writing although they may have struggled with communication, while others may have been more fluent in speaking but adopted different forms of English. Students who are confident and extrovert may be more prepared to use the new language and make mistakes, but that does not mean they are proficient. Proficiency must surely include accuracy and accuracy must be assessed by standardised criteria. There needs to be some form of standardisation



when making judgements on performance; if there is no standard form of English to use for assessment, this must affect the reliability and validity of any testing system.

**Financial background** also emerged from the Stage 3 interviews as an important factor, in particular when this resulted in increased opportunities to practice English both in school and when travelling abroad. Stage 3 Participant 1 commented that ‘a better financial background means the ability to attend private schools and to travel, so they are exposed to languages’. The private school system is subjected to demands from parents for a higher quality of education, and therefore resources are often greater. However, not only the materials but also the teachers were likely to be better qualified for teaching English. Teachers in the state school system are all Saudi and, consequently, their own level of English may be limited as they have gone through the same school system. However, in private schools, there are not the same limitations on employing foreign staff and thus some of the teachers may be native speakers of English or may have studied in English-speaking countries.

Even teachers from other Arab countries, let’s say Syria, Egypt, or Lebanon, she may have gone to an English school, or a French school, her education may have come from a private school or an international school in her own country and she has a better proficiency than a Saudi teacher (P.1).

Students from the private school system are thus less likely to be placed in lower level classes. Some may also be exempted from the English language course altogether, as they have taken a TOEFL exam or IELTS test instead. It is clear that students who come from a family background, who can afford private education, have an advantage. Private schools start teaching English at an earlier age, and they provide intensive language courses. This gives such students an advantage and they will normally go straight into the English course at an advanced level. In addition, private and international schools; often teach all subjects through the medium of English and this gives extra confidence to privately-educated students (Participant 6).

Being able to use smart phones and computer programmes to enhance their English can also help those from better financial backgrounds, according to Participant 4. This means such students have more opportunities to practise the language. Because of this extra practice outside the classroom, students can improve their skills and gain better results in tests.

Additionally, some wealthier students have an added advantage of frequently speaking English in the home (Participant 6) and may be more familiar with the topics covered in the OOPT or ELI modules. However, Participant 4 did not believe that tests reflect the actual level of English proficiency for language users not their success in the future; she was in fact opposed to all kinds of testing and believed that these are outdated methods of assessment.

Other factors, which emerged from the analysis as having implications for students' achievement, were their **health and wellbeing** as well as the way students responded to certain teaching approaches and their own ability to learn languages. In interviews, some teachers suggested that the health of the student may adversely affect performance, especially with English being a foreign language and the other subjects all being easier, as they were taught in Arabic. A student weakened by ill health 'may need more focus and more strength to absorb the language' (Participant 2). Yet it is probable that ill health only played a role in exceptional cases and can therefore only be viewed as relevant for a small number of students. Nonetheless health-related issues can affect an individual's achievement; these may be either physical health or mental health issues, such as anxiety, or depression that may lead to a weak performance. Mental wellbeing may be affected by family problems, as Participant 5 suggested, and this may influence the student's overall proficiency.

The relevance of teaching as well as the way students responded to particular teaching approaches was also found to influence achievement. As Participant 5 argued, if students are taught one hundred words of vocabulary in a module, then these words should be ones they will actually use. Participant 5 pointed out that the unit on shopping is always well received probably because students perceive it as useful. They may go to restaurants where the waiter is not a Saudi; he may be from the Philippines or from Sri Lanka, and does not speak Arabic. However, this waiter will probably be able to speak English, so there is an opportunity for a Saudi student to practise the language. In Saudi Arabia, there are many other nationals working in the Kingdom and they come from countries like India or other parts of Asia, where English is spoken; even in the shopping centres P.5 believes there are opportunities for using English. If more emphasis were to be placed on such opportunities, the learning of English may have more relevance and students may actually enjoy practising their language skills.

### 5.4.3. Drawbacks of the modular system

The modular system continued to emerge from the interviews as problematic. Firstly, it does not follow the university's academic calendar, which adopts a **semester system**, resulting in **conflict with other courses** across the university. Secondly, the English language programme **offers four levels** of competence instead of the six levels drawn from the CEFR used in the OOPT. Drawbacks of the modular system were identified also with regards to **contact hours** as teachers reported that students felt the courses were too short to enable them to make adequate progress in the language. A further drawback concerns the teacher-student rapport. Some teachers felt that the constant change in lecturers did not allow teachers enough time to get to know the students' weaknesses and help them develop their language skills accordingly. Participant 5 argued that the **full semester of learning** gives more confidence to students, as they are not **moved** on to new classes and new teachers all the time. This helps students' development as time is not wasted in adjusting to new teaching approaches and different peer groups. According to P.5, the decision to adopt a modular system was one man's decision and he wanted the whole university to use a modular system in the foundation year. Only the English language department agreed to this, and it has now gone through an evaluation and accreditation process; if anything were to be changed now, it would require a major change and become an immense procedure. Nevertheless, there are concerns that the modules do not provide enough time for students to make adequate progress. Participant 2's suggestion was to keep students with the same teacher for three months, so they could **build a rapport** and chart progress. In the current system, teaching takes place just for six weeks and then students are assessed and move on to the next course. In this regard, Participant 2 commented that this approach 'causes confusion' as the students 'have too many teachers'. It seems that this lack of continuity is a problem for teachers as well as students, as they do not get to know their students and cannot teach them according to their needs. According to Participant 3, this means they do have time to consolidate their learning. They want to finish the course as quickly as possible in order to progress to the next level, but they miss a lot of information. The amount of learning that goes on in five or six weeks is limited.

Participant 7 commented the students who have passed all the required modules to reach Level 3 ‘haven’t actually achieved level 3 in the language’. She added that ‘if they took it again, the OOPT, they will not be placed in level 3, because they are still stuck in level 1. They are only on level 3 because they passed level 1 and level 2’. As Participant 7 argues, they are estimated to be at Level 3 simply because they passed the lower level modules, not because their language ability has significantly improved. It is therefore an area of concern, as the English programme becomes ‘a race to the finish’, according to P.7. It becomes a measure of quantity, not quality.

If it continues like this, it will be a complete failure as it is not facilitating at all. They have to do one of two things: either change the quarterly system to just one semester and give teachers time to finish the book and practise; or keep the quarterly system and teach just half a book. This would give students a chance to absorb everything and have enough time to practise. It may take two years but why not? If you are looking at competences and want them to be really good, then in two years they will have time to practise writing and speaking and perhaps write an essay. (P.7)

The OOPT may also have a greater chance of being perceived as successful, if the students are not being forced through a system that expects so much from both teachers and students. As P.7 suggests, students need time to consolidate their learning and rushing through the books, with the mere aim of completing them within a short period of time, does not actively support learning.

Participant 8, however, argued that there is an inherent issue with teachers’ own attitude towards the profession. She argued that they do not support their students in gaining proficiency. She felt that there were many teachers who have no sense of vocation; they simply see teaching as a job. These teachers come to class without preparing any materials and no proper lesson planning as they just use the book.

For example, in each class you have multi-level students who need different attention. After a week of observing my students I get an idea about their levels and my main concern will be designing tasks for each group of students. But these teachers are not a guide for their students, they simply spoon-feed them and do not give the students the opportunity to think or be creative. On the other hand, there are some teachers who deliver the information in very interesting ways with worksheets that they work through with their students and you can definitely see better results. (P8)

Participant 8 makes reference to the more student-centred approach to teaching, more commonly observed in Western education, where the student actively participates in the learning experience. The teacher provides students with tasks and resources to engage them

and thus the student becomes more motivated and willing to participate. From the interviews, however, it appears that the majority of teachers lean heavily on the textbook and adopt primarily a teacher-centred approach. This is an important finding and will be further discussed in the following chapter.

## **5.5. Summary**

The chapter provides an overview of the key issues as they emerged from the questionnaire and interview analysis in relation to teachers' beliefs about the OOPT, the modular system and students in the ELI. The initial interviews (Stage 1) provided a broader perspective on teacher's views on the OOPT and also served to inform the questionnaire design, while the second set of interviews (Stage 3) took place after the questionnaire and therefore the participants had a better view and understanding of the research aims. Stage 3 participants were also given some sample OOPT questions to look at so, even if there were not familiar with the OOPT, they could at least provide some comments about the extract.

The initial interviews exposed the lack of familiarity with the OOPT among many of the participants. This information was invaluable as it informed the analysis of the questionnaire and the writing up of the interview schedule employed in Stage 3. For instance, there were items on the questionnaire, which referred specifically to the OOPT, such as question 1 (the OOPT as an effective tool) and question 2 (the OOPT reflects students' language ability). When looking at the results obtained from all participants, the mean scores indicated that the average response was "I don't know". However, based upon the responses provided in the initial interviews, I decided to carry out a Mann-Whitney test to explore whether responses would be different if I separated the participants who were familiar with the OOPT from those who were unfamiliar. The Mann-Whitney test exposed the difference in responses for items one and two between the two groups, thus confirming that it was the unfamiliar group which could not comment on these two items.

Furthermore, the stage 3 interviews served to better illustrate the findings of the previous two data collection stages, for instance by illustrating reasons why many teachers were unsatisfied with the modular system. The qualitative and quantitative data similarly highlighted the fact that the majority of teachers were not familiar with the OOPT, so there was little they could say about it. Among those who were familiar with the test, some saw it as an effective tool,

while others felt that the OOPT results often led to misplacement issues with students being placed in levels too low or too high for their ability. Some teachers, therefore, believed that the OOPT results did not always reflect students' actual proficiency in the language. Both the questionnaire and interview data also portray the perceived difference in ability between Science and Arts students as well as pointing out issues with the modular system. The following chapter explores some of the findings in greater depth making cross-reference with the literature, in order to frame and contextualise teachers' beliefs on the OOPT and the students at KAU.

## CHAPTER 6 DISCUSSION

### 6.1. Introduction

The chapter discusses the findings drawn both from the questionnaire and interview data in relation to the four research questions, which guided the study. Findings which provided evidence in support of the first two research questions were subdivided into three categories, namely: perceptions, uses and consequences of the OOPT. These subcategories draw upon Messick's (1989) unified view of validity, as previously described in chapter 4. There were also some findings which were related to research question 1 but were not relevant to any of these subcategories. These appear under the section "other" in the summary of key findings (see below). Key findings in support of research questions 3 and 4 are also summarised in the table and these refer to teachers' views on factors affecting students' achievement and areas for improvement within the ELI.

**Table 8. Summary of key findings**

RESEARCH QUESTIONS	KEY FINDINGS
<b>1. How do teachers view the OOPT?</b>	<ul style="list-style-type: none"><li>• Teachers who were familiar with the OOPT viewed it as an effective tool and felt that it reflected students' language ability.</li><li>• Few teachers viewed the international nature of the OOPT as an advantage, while others criticised the OOPT because they felt the cultural content made the test unfair for Saudi students.</li><li>• Computer literacy and preparation were regarded as linked to OOPT results.</li></ul>

	<ul style="list-style-type: none"> <li>• OOPT test scores were used to place students into the 4 ELI modules – this is different from the recommendations of the test developers</li> <li>• Some teachers referred to the mismatch between placement test and curriculum</li> <li>• Some teachers described the difference between the 6 CEFR levels of the OOPT and the 4 CEFR levels adopted by the ELI as problematic for placing students correctly.</li> <li>• Most teachers were unfamiliar with the content of the OOPT.</li> <li>• Some teachers appeared not to understand the true purpose of the OOPT and described it as an achievement test (e.g. by comparing it to tests like the TOEFL and IELTS).</li> </ul>
<b>2. How does the OOPT impact on placement and teaching, in the opinions of teachers?</b>	<ul style="list-style-type: none"> <li>• Some teachers felt that OOPT results led to misplacement into levels too high or too low i.e. they did not reflect students' true ability.</li> <li>• Misplacement can lead to diverse ability within the same course. Some teachers viewed this as limiting the extent to which they could support students' learning.</li> </ul>
<b>3. What are the factors that might affect students' achievement in both the OOPT and during the English module, in the opinion of teachers?</b>	<ul style="list-style-type: none"> <li>• Students' motivation (both intrinsic and extrinsic) towards studying English.</li> <li>• Affective factors (e.g. stress, anxiety)</li> <li>• Educational factors - Teachers felt that students who attended private schools had more advanced language competence and generally did better in the OOPT and in the</li> </ul>



	<p>ELI modules. They also felt that students who had learnt English from an early age were better prepared.</p> <ul style="list-style-type: none"> <li>• Socio-economic factors were also highlighted as being linked to students' previous learning experiences (e.g. some students had visited English speaking countries).</li> <li>• Attitude towards language emerged from the interviews as a factor affecting students' motivation for studying English.</li> <li>• Specialism – Teachers also felt that Science students were generally better prepared compared to Arts students.</li> </ul>
<p><b>4. How can the system be improved to support students' achievement?</b></p>	<ul style="list-style-type: none"> <li>• Teachers suggested redesigning the curriculum so that modules last for one semester instead of six weeks</li> <li>• It was also suggested that the ELI could assign the same teacher and partner teacher to a semester-long module sequence (e.g. level 1 and level 2).</li> <li>• Some teachers mentioned the use of in-house placements tests designed to match the Cambridge Special edition books and the four CEFR levels in the ELI as a better alternative to the OOPT</li> <li>• If the ELI wants to continue using the OOPT, teachers need to become more familiar with the content so that they can fully understand what skills their students need to develop and how the university can ensure students are fully informed of the OOPT before they take the placement test.</li> <li>• It seems to be that there is a need for in-service work with teachers so that they understand better the various purposes of assessment. A number of the teachers' comments appear to reveal a lack of understanding.</li> </ul>

The key findings summarised above, are discussed in detail in this chapter with reference to the literature. While chapter 5 reports the results of the investigation in relation to the themes,

sub-themes, codes and sub-codes which guided the data analysis, this chapter discusses key findings in relation to the research questions. Furthermore, there are findings which emerged from the analysis, which do not specifically answer any of the research questions. Some of the emerging findings add to our understanding of the complexities behind administering placement tests within this unique cultural context, while others have implications for theory, in particular with regards to the current literature on placement tests.

With regards to the OOPT, in some instances, teachers held contradictory views. For example, some teachers saw the international nature of the OOPT as an advantage, while others felt that it was not relevant to Saudi students because of its cultural content. Some teachers also felt that the OOPT results did not reflect students' actual language ability and therefore were misplaced in an incorrect ELI module. Teachers further made reference to factors, which could affect students' performance, both in the OOPT and in the ELI modules. These include both educational factors (e.g. their school experience) as well as affective factors (e.g. stress and anxiety). A key factor was also student motivation, which emerged several times in interviews and in the questionnaire in relation to students' specialisation and previous learning experience.

## **6.2. Validity of the OOPT for ELI students at KAU**

As previously mentioned, the majority of teachers interviewed were unfamiliar with the OOPT and, therefore, could not comment on the test. Those who were familiar with the test expressed different views. Some highlighted the advantages of employing the OOPT as a placement test, while others made reference to its drawbacks. Teachers' lack of familiarity emerged as a key finding, which exposes the extent to which teachers are not provided with sufficient training and guidance concerning the OOPT.

The findings which provide evidence in support of the first two research questions can also shed light on the validity of the OOPT, according to Messick's unified view of validity, as teachers commented on their perceptions of the test, its use within the ELI and consequences of employing the placement test within this context.

With regards to perceptions of the OOPT, teachers who were familiar with the test generally agreed that the test was an effective tool and it generally placed students into the appropriate level. They also agreed that the OOPT reflected students' abilities in the English Language. Teachers, however, had different viewpoints on the cultural relevance of the content for Saudi students, thus questioning the fairness of the test for ELI students.

The findings also help illustrate the face validity of the OOPT, through teachers' experiences and views of using the placement test within the ELI. As previously mentioned, face validity refers to the test's 'surface credibility or public acceptability' (Alderson, Clapham and Wall, 1995, p. 172). Face validity is therefore, concerned with how students, administrators or non-expert people view the value of the test. Gipps (2012) points out that 'the responsibility for valid test use is put onto the user not the test developer (p. 50)'. This is an important point as it may be that some of the issues attributed to the OOPT itself, as they emerged from the interviews, were actually more strongly related to the way the OOPT was used, the lack of preparation on behalf of students, the lack of familiarity teachers had with the test or the way the OOPT results were employed for placement purposes at the ELI. Furthermore, as Messick (1989) claimed, we cannot assume that teachers share the same frame of reference as the test developer, and it should be acknowledged that test use cannot be considered independently of its local context. The test developer may have envisioned a certain target population and may have designed the test accordingly, however the test takers, and those administering the test, may have very diverse learning experiences.

### 6.2.1. Limited understanding of the OOPT

The data analysis exposed a limited understanding of the OOPT on behalf of teachers. Some described the OOPT as an achievement test or used the term placement test but referred to it as if it had been designed to assess test-takers' overall competence in the language. It emerged, therefore, that teachers were unclear about the purpose of placement tests and what they were designed to measure. It was evident that staff had little knowledge of the characteristics of the OOPT in particular and how it was being employed in their context, i.e. in the ELI at KAU.

The OOPT was mainly designed by its developers to facilitate placement into an appropriate English language course at the institution where the students would study. It therefore, was not designed to provide a comprehensive picture of students' communicative competence in English across the four skills but rather to simply group students into different levels to facilitate teaching. However, while placement tests are not intended to fully assess competence in the language, they do not function in a vacuum. Whether or not a placement test 'works' does not simply relate to the quality of the test itself, but also to the ways in which results are used. As previously mentioned, in the ELI students are placed in an English language module as a result of their OOPT scores, however the marks range for the levels are different from the ones recommended by the OOPT test developers. The ELI also offers only four different levels instead of the six suggested by the publisher. Hence, in this context, the results are not used as recommended by the test developers and it is possible that the marks range adopted in the ELI may be linked to some of the misplacement issues reported in the teacher interviews. Purpura (2007), in a research paper describing the meaning of OOPT scores, argues that regardless of how well test developers have designed the test, they cannot prevent people from misusing it. If this takes place, then 'invalid use' is made of a 'valid test'. The OOPT, he suggests, has been designed to serve one primary purpose: to place students accurately into an English language class appropriate for their needs ([www.oxfordenglishtesting.com](http://www.oxfordenglishtesting.com)).

While proficiency test scores have been used for placement purposes in some institutions, it is important to understand that they have not been designed for this purpose. Fox (2009)

found that using proficiency test scores, such as IELTS and TOEFL, for placement into English for Academic Purposes (EAP) courses was problematic. Students placed in the same level, had very diverse abilities and there were huge differences between those at the top and those at the bottom of the EAP classes (Fox, 2009). While teachers at the ELI also complained about significant differences in ability within one level, the ELI *does not* employ IELTS or TOEFL scores for placement purposes. The only test used for placement purposes at the ELI is the OOPT, which was designed for this specific purpose. While there are some similarities between proficiency tests and placement tests, one of the observable differences between them is that placement tests are not designed to inform teaching and learning but merely to place students in an appropriate level.

A crucial question, which emerged both from the literature and the data analysis, concerns whether misplacement causes significant disruptions in teaching and learning. This reflects Hughes' (2003) argument, where he claimed that placement tests are designed to provide 'an indication' of overall ability rather than determining a student's proficiency level or competence across the four skills. Furthermore, he argued that placement tests are simply used for making decisions regarding which course is most appropriate for students. Therefore, these decisions are 'not of critical importance' and since people do not generally prepare for placement tests, 'there is no need to worry about possible backwash' (p.186). Hughes (2003) additionally argues that where students are misplaced by the placement test, 'they can usually easily be moved to a more appropriate class', however, while moving students to a higher or lower level may appear easy, one should also consider the social context (p.186).

Through careful analysis of the interview data, it also became apparent that teachers spoke differently about students who scored high in the OOPT compared to those who scored low. While they did not express clear criticism of students who did not perform well, some of the assumptions appear judgemental. One teacher commented that students in level 1 and 2 have weak grammar, vocabulary and speaking. She added that she thought students had guessed answers in the OOPT and did not know what they were doing. She argued that there is something wrong with students in level 1 and 2 and the placement test is not placing students correctly because their English is extremely weak. This teacher appears to believe that

students were misplaced in a higher level, presumably in level 2 instead of level 1, as a result of having guessed answers. The teacher's lack of trust in the students and criticism of their competence is an example of how lower level students may become stigmatised as weak and not hard working. Some teachers also held strong views on students who were placed in higher levels as a result of the OOPT. One of the teachers interviewed reported that you can tell that students placed in levels 3 and 4 have travelled abroad and come from families with higher SES. Here, the teacher is making a generalisation about students in levels 3 and 4. These students are viewed in a positive light, which once again, suggests that some teachers infer all kinds of information from the test results students obtain in the OOPT and the level in which they are placed. Sanford (2005) argues that, 'unexamined stereotypes shape teachers' expectations of students' limited opportunities for them to explore and define alternative realities' (p.306). Some teachers thus base their judgements on the observation of certain traits observable in weaker or stronger students, without exploring alternative factors, which could also affect test performance.

Concerning the purpose of the OOPT, one teacher fully understood that the aim of the test was to help place students in the correct level and not to provide a comprehensive picture of their proficiency in English. She felt that the OOPT was generally effective at placing students correctly and therefore found the test 'valid' and 'serving its purpose'. She also acknowledged that the test 'doesn't test everything' and therefore it 'doesn't reflect students' competence in English language'. Another teacher, who had taught at KAU before the modular system was introduced, commented that previously she had to teach students with a very wide ability range and had to modify the Cambridge textbook to serve the needs of all the different levels. Others also agreed that the OOPT was generally effective in placing students into the correct level and pointed out that most students placed in level 1 had very similar levels and language background.

Some teachers, however, felt that the placement test should match the content of the ELI courses. One of the participants argued that such tests should be linked, in a way, to the curriculum. She felt that there was a mismatch between the Cambridge books employed by the ELI and the OOPT and commented that the ELI had recently decided to use Cambridge books instead of Oxford for the teaching of ELI courses and that this would certainly affect

students' achievement. While it may be preferable, in order to ensure consistency, to adopt materials from the same publisher, the fact that some teacher held very strong feelings about the apparent mismatch between the OOPT and the curriculum appears to once again highlight a confusion about the purpose of placement tests.

### **6.2.2. Cultural content of the OOPT**

The cultural content of the OOPT repeatedly emerged from interviews as an area of concern. Some teachers appreciated the international nature of the test and commented that, when it comes to language testing, the content should be the same everywhere to avoid bias. However, others felt that the OOPT should have a special edition, as is the case with the textbook employed by the ELI, which is more culturally relevant to Saudi students. For instance, one of the participants pointed out that female Saudi students were very unlikely to have any knowledge of Rugby, as women in Saudi Arabia do not practise sport. It was further pointed out, that the test should be culturally sensitive by avoiding topics, which make reference to drinking alcohol or homosexuality, which are taboo in Saudi Arabia. As a result of the cultural difference, it is unlikely that students would have much prior knowledge of these topics and may therefore, not fully understand the interactions. Gipps (2012) suggests that where assessments are unfair to one particular group or another, the test may be considered biased. In this case, if the content of the test predominantly reflects a Western cultural lifestyle, it may be viewed as unfair to students who are unfamiliar with the social practices in the given context.

However, Sulaimani and Elyas (2018), who explored the *New Headway special edition*, which was previously used in the ELI, found that the Special Edition series 'repeatedly excluded females' images, highly segregated male and female images, limiting females to family and shopping contexts and covered females' heads with various hats' (p.72). They concluded that the series presented 'a biased gender representation rather than a careful representation of gender that meets the needs of the Saudi context' (ibid.). Therefore, adapting a textbook or test to make it more culturally sensitive may also result in bias or misrepresentation of a particular ethnic group or gender, as described above.

Bias due to the content of an assessment may be regarded as ‘unfair penalisation’ McMillan (2008). In such cases, an assessment may be more challenging for students unfamiliar with the content compared to those who are familiar with it. As previously mentioned, bias is present when one group is disadvantaged as a result of their gender, race, language or other characteristics. Where the content of the test is unrelated to what students have learnt in their educational setting, students who are unfamiliar with it may find the test more challenging and underperform as a result. McMillan (2008, p. 60) provides a good example of this:

Suppose you take a test that measures your aptitude by using mostly rural, farm-oriented content. The items covered touch topics as types of cows and pigs, farm equipment and winter crops. If you grew up in a city or suburb, do you think your score will be as high as the one of those who grew up on a farm?

McMillan (2008) also suggests an example of unfair penalisation, where boys might be advantaged compared to girls. Coincidentally, the example refers to the content of sports. He questioned whether test items containing sport content may give boys an unfair advantage. As with the example of farming, ‘in each case membership in a particular group or background unrelated to instruction influences the score’ (McMillan, 2008, p. 60-61). However, while it may appear obvious that knowledge of the content or cultural context presented in a test may facilitate comprehension, Maddox (2014) seems to suggest otherwise. The study explored the dynamics of literacy assessments in a rural area in Mongolia. The findings highlight some of the problems that nomadic herders faced when confronted with realistic textual knowledge on camel herding. Maddox (2014) found that while ‘the local content increased people’s motivation and interest in the questions (...) the test item was ‘too real’, since local knowledge acted as an unexpected distractor’ (p.487). Maddox (2014) therefore concluded that people who are less familiar with camels or the geography of Mongolia may have been more likely to obtain their answers exclusively from the textual information as opposed to referring to their own knowledge about the subject.

Thus, while test developers should be cautious of unfair penalisation as a result of test content, it is practically impossible to remove all types of bias from such tests. Questions need to be written in ‘a content’ and invariably this content will unfairly penalise students to some extent (ibid). Rauf (2015) also pointed out that the topical content of achievement tests should be



related to the topics students have already studied. While this concerns achievement tests, it would be preferable if placement tests would take into consideration topics, which students may have covered in secondary school.

While it may be argued that when one learns a language one also learns about culture, because of the teaching approach adopted in Saudi schools, it is unlikely that much focus would have been placed on learning about the target culture. Al-Seghayer (2014) suggests that students in Saudi schools are not taught the techniques to enable them to interpret spoken and written forms of English. Nor are they taught ‘when to say what to whom’ (p. 22). Teaching is largely confined to ‘extensive drills of grammar rules and the monotonous repetition of words and phrases’ (ibid., p. 22). As a result, the majority of students, in particular those who attended public schools, come to university with little experience of using English for communicative purposes or knowledge of the target culture. The lack of familiarity with the target culture has been previously found to affect ways in which test items are answered. Abu-Rabia (1998) found that Arab students learning Hebrew in Israel were more successful on reading tasks in which the content was familiar to their Arab culture compared to tasks with unfamiliar cultural content.

However, determining whether a test is biased or whether the test takers have a different underlying level of attainment is extremely difficult (Gipps, 2012, p.125). It is therefore difficult to establish whether low achievement in the OOPT may have been linked to the cultural content or to other factors, such as test design, students’ lack of preparation or poor language skills. One of the teachers interviewed argued that the content of the OOPT should be more culturally relevant to Saudi students. This does not mean ‘changing names such as George to Abdullah’, but rather rethinking about the topics. For instance, she pointed out that female students, in particular, would probably not be familiar with sports. She further pointed out that ‘foreign metaphors’ should be removed from the test, as students would not be familiar with them. As previously mentioned in Chapter 3, Brooke et al. (2012) found that the OOPT covered a wide range of cultural situations that Arab students were unlikely to have encountered. It is difficult to determine to what extent the occurrence of idiomatic expressions may affect students’ ability to comprehend items on the test, in particular because

the OOPT is a CAT, however the teacher raises a valid criticism, which could be explored further.

A further issue, which emerged in relation to the cultural content of the OOPT, concerned the presentation of interactions between native speakers and the focus on cultural references specifically related to English-speaking countries (in particular the use of British English in the UK). This is interesting as it introduces the dimension of world or international English. Global English is perceived as belonging to non-native speakers and national standards of the English language should not be taken as the model, according to Brumfit (2001). The belief that the native speaker provides the sole source of reliable linguistic information was drawn from Chomsky (1965) and has long been discredited. Byram (2009) for instance argues that the ‘intercultural speaker’ offers a better model than the native speaker. However, this does not have full support from other researchers; a study of Chinese students found they preferred native speaker models (Pei and Xing, 2016). In addition, students going through Saudi universities now gain scholarships to study abroad. This often entails attending British, American, Canadian or Australian universities, where they will need to deal with standard English models. It is therefore more likely that they will benefit from internationally recognised tests, such as the OOPT, since this is the type of English they will mostly need to succeed academically.

### **6.2.3. Skills tested in the OOPT**

The reliance of the OOPT on testing exclusively the receptive skills of reading and listening was criticised by some teachers, as they felt that the test could only assess the extent to which students understood language but failed to assess the productive skills of speaking and writing. While some aspects of writing and speaking can be assessed indirectly (for instance the OOPT indirectly tests grammatical knowledge through the reading comprehension exercises), these often only provide a limited and incomplete picture in relation to the range of knowledge and abilities involved in language production and interaction.

In interviews, teachers expressed mixed views on whether all four skills should be included in the placement test. Some related fairness to the skills tested in the OOPT and thus felt that

the OOPT should test all four skills in order to fairly measure students' ability. They argued that since all four skills are taught in the curriculum, they should also form part of the placement test. Other teachers, however, argued that university applicants had such weak competence in speaking and writing that it would be unfair to assess these skills in a placement test. Another teacher commented that it would be time consuming to test all four skills and felt that this was not necessary since the placement test works as 'an elimination process' to 'just decide who stays and who goes'. Another participant commented that she did not think it was essential to test all four skills, because Saudi students are generally very weak in speaking and writing, since there has been very little focus on these two skills in public schools. She added that, as a result, it might be unfair to include speaking or writing questions in the placement test, as these would not reflect students' ability.

Teachers' diverse views on whether the OOPT should test all four skills raise an important point, however also expose a possible lack of understanding of the purpose of the OOPT. Morante (1989), Scott-Clayton (2012) and others have strongly advised not to base judgment of students' levels on a placement test score, especially if it does not integrate all four language skills. They recommended that practitioners use multiple methods in addition to the obtained score to make decisions regarding students' actual levels of English language proficiency. There are hence authors, who agree that placement tests should integrate all four skills. Hughes (2003), however, points out that when testing is carried out on a very large scale, test developers and teachers do not want to wait to receive students' marks. The scoring of tens of thousands of compositions hence does not seem to be practical proposition, although it may potentially result in greater accuracy. However, simplifying the marking process (for instance by employing multiple-choice questions, which can be marked by computers) has not always given testing a good name! (Hughes, 2003, p.3). One skill, which is often assessed in placement tests, is test takers' knowledge of grammar (Green, 2012). Grammatical knowledge can easily be tested through multiple choice or gap-fill activities, which can also be marked electronically. Hence, while there may be advantages to testing all four skills, for practical reasons, many placement tests do not.

Hughes (2003) argues that whether or not grammar is central to an institution's teaching, students' grammatical ability, or lack of it, sets limits to what can be achieved in terms of

performance. Successful academic writing, for instance, depends, to some extent, on the students' command of grammatical structures. Therefore, 'there appears to be room for a grammar component in at least some placement tests' (Hughes, 2003, p.173). While placement tests, which rely heavily on grammar, may be criticised for their lack of authenticity, it would be interesting to explore whether these may be more culturally appropriate, given that students in Saudi Arabia often lack knowledge of the target culture, and have not had much practice in listening comprehension.

#### **6.2.4. Computer literacy and administration of the test**

Teachers interviewed also made reference to the students' lack of familiarity with computers and computer-based testing. They felt that, for some students, the poor computer literacy skills may have affected their performance in the OOPT. The relationship between individual characteristics, knowledge and skills of test takers and test performance has similarly been identified in the literature. Gipps (2012, p.126) argued that differential performance may be due to factors in the test itself, since some individuals may have an equal level of knowledge or skill but be unable to demonstrate their full potential because of their lack of familiarity with a characteristic of the test. This might be because of the language used, or artefacts in the illustrations or text, make the task more meaningful to one group but less meaningful to another. Purpura (2007) claims that the OOPT measures test takers' communicative language ability in order for the exam scores to be used relatively accurately in placement decisions for language programmes that are aligned with the CEFR ([www.oxfordenglishtesting.com](http://www.oxfordenglishtesting.com)). Differences in performance between students, however, could also be linked to students' previous learning experiences. While some may have been taught English according to a communicative teaching method, others may have focused primarily on grammar and vocabulary.

#### **6.3. Teachers' beliefs and attitudes**

The data collected in this study can be understood as subjective, since the descriptions of students and possible issues related to the OOPT need to be viewed from the perspective of the participants' interviews. In some cases, teachers provided generalisations of issues

relating to the placement test or regarding factors affecting students' achievement, but it should be highlighted that this cannot be interpreted as an objective reality. This section therefore, seeks to discuss teachers' beliefs and attitudes on the OOPT, the modular system and students in the ELI while highlighting that even what teachers present as facts, may not represent an objective reality. It should also be noted that beliefs play an important role in many aspects of teaching and life, and that beliefs influence how information is perceived and 'colour memories' in order to 'frame our understanding of events' (Borg, 2001).

The questionnaire highlighted stronger views on some of the questions compared to others, indicating that teachers were less unsure about their views on certain items. For example, teachers held strong views on question 9 (science students perform better than those majoring in the Arts), and on question 7 (students from private schools perform better in English compared to those from public schools). They were also in agreement on question 14 (learning English from an early age) and question 15 (the role of motivation).

Unlike other countries, where the choice of specialisation is not so strongly related to students' ability, in Saudi Arabia students are advised to specialise in the Sciences as early as high school, if they have good results, while they are only offered the option of studying the Arts, if they have lower grades. University admission reflects the same view of the two specialisations and thus, culturally, Science students are viewed as more academic and hard working. It has been previously pointed out that in contrast to the science track, the arts track choice is not encouraged. Furthermore, high school graduates of the science track generally have better HE and employment opportunities. Graduates of the Arts track, cannot enter prestigious university faculties such as medicine or engineering. Additionally, students who lack confidence in their academic ability may select the arts track to avoid more demanding subjects, which are compulsory in the science track (Al-Jabri, 2003).

Among teachers who were familiar with the OOPT, many also expressed strong views with regards to the cultural appropriateness of the OOPT's content for Saudi students. While some viewed the international nature of the test as an advantage, a number of teachers felt that the content was not culturally relevant. This point is further discussed in the following section in relation to teachers' views on the validity of the OOPT for ELI students at KAU. The

section, furthermore, reviews teachers' perspectives on the validity of the OOPT with reference to what they said with regards to the familiarity, fairness and effectiveness of the test in relation to the purpose for which it is employed in the ELI.

#### **6.4. Factors affecting students' test performance**

The analysis highlighted a number of factors affecting students' performance, in the opinion of teachers, both in the OOPT and in the ELI modules more generally. These included characteristics linked to students' health and wellbeing as well as previous educational experience. O'Sullivan and Green (2011) distinguish between three types of test taker characteristics: 1. Physical / Physiological 2. Psychological 3. Experiential. This categorisation informed the classification of the different factors (see chapter 3).

O'Sullivan and Green's grouping of the different kinds of test taker characteristics are helpful in interpreting the findings. The interview data, in particular, exposed students' previous learning experience and motivation as key factors, which teachers regarded as influencing performance in the OOPT and in the ELI modules. The questionnaire results also highlighted the importance of motivation and learning English from an early age.

##### **6.4.1. Previous learning experience**

Students' previous learning experience emerged from the interviews as a key factor affecting students' achievement in the OOPT, in the ELI modules and influencing their overall attitude towards learning English. Previous learning experience consists of a number of factors, including the type of school students attended, their previous experience of learning English and the age when they began learning English. In interviews, teachers also felt that socio-economic background was often associated with the kind of school students attended and their opportunities to travel and to be exposed to English.

Generally, teachers felt that students who had attended public secondary schools were not familiar with communicative approaches towards language learning as they had been taught mostly grammar and reading. Their previous experience may have influenced their

expectations of studying English at university. Puchta (1999) argued that beliefs are guiding principles for students' behaviour and strong perceptual filters and act as if they were true. Student may, for example, hold the belief that their grammar is weak and that they need more practice, because this is what their teachers told them, yet fail to recognise that their listening and speaking skills are far weaker.

Likewise, teachers may also hold strong beliefs about what their students should learn. Kennedy (1996) hypothesises that real and effective change in teachers' practices can only occur through a change in their beliefs because beliefs shape the way teachers behave. One of the problems with the teaching of English in schools, concerns teachers' subject knowledge. School teachers' knowledge of English is often weak and, therefore, they feel more comfortable simply encouraging their students to memorise vocabulary and structures in order to pass the exams. Another teacher suggested that school teaching still adopts very traditional approaches, because English teachers are not adequately qualified and suggested that they complete professional developments courses to update their skills. The result, however, is that students leave secondary education not only with poor competency in English, but also with a negative attitude towards the subject and low self-confidence about their ability to succeed. While a number of teachers criticised the teaching of English in public schools and generally agreed that students coming from private schools were better prepared, one teacher related her personal experience, which highlights that this is not always the case. She gave the example of two of her relatives: one had attended a public school, while the other one had attended a private school. Both took the OOPT upon entering KAU and, unexpectedly, the one who had attended a public school was placed in level four, while the one who had attended a private school was placed in level two.

#### **6.4.1.1. Equity between financial advantage and better opportunities**

Socio-economic factors emerged from the interviews as an important factor. The majority of teachers, however, appeared to equate higher financial status with an increased opportunity to attend private schools, learn English from an earlier age and travel to English speaking countries. In fact, 15 out of the 18 teachers interviewed agreed that financial background affected students' success in the ELI modules, an overwhelming majority. However, while they were asked about financial background, their comments related predominantly to the

opportunity to attend a private school and to travel. From this, teachers inferred that these students must have had the financial means to do so. Surprisingly, this is one finding in which the quantitative and qualitative data are in disagreement. In the survey results, the mean score for question 10 was 2.84, situating responses between “disagree” and “don’t know”. On the other hand, the majority of survey responses agreed to question 11, which asked whether teachers felt that travelling abroad and going to private schools would enhance students’ language competence. The mean scores for question 11 were situated between “agree” and “strongly agree”, with a mean value of 4.35 thus confirming the interview data. This is an interesting finding, as it would seem to suggest that, in interviews, teachers did not view the two questions as inherently different but rather saw financial advantage as a predictor for better opportunities, with only some teachers who recognised that success may be linked to other factors. It is also possible that the difference in the demographics of the survey respondents and of the teachers interviewed affected the way item 10 in the questionnaire was understood since most of the teachers interviewed were Saudi citizens. However, two of the teachers interviewed were not Saudi and their views did not differ from those of the other respondents. What is interesting to highlight is that, in the questionnaire, participants clearly interpreted items 10 and 11 as distinct from each other, while in interviews teachers spoke about financial background in relation to increased opportunities to travel and to attend a private school. Furthermore, the difference in responses obtained from the interviews may have been related to social desirability bias. In interviews, participants may have assumed that previous school experience was to be understood as linked to higher SES.

As highlighted above, in interviews teachers generally made reference to students’ SES with reference to the opportunities that some students may have had compared to others. Some teachers were of the opinion that students who had attended private schools and had had the opportunity to travel to English-speaking countries often had better language skills and more positive attitudes towards learning English. Fan (2011) suggested that SES does not only influence language learning but also it affects learners’ self-related beliefs and motivation to learn, which in turn can affect achievement. In this regard, one teacher commented that students coming from private schools have excellent English skills and this affects their views about English. These students also often have the opportunity to travel abroad and therefore



have experience of the outside world. Another teacher felt that students who could not visit English-speaking countries or attend private schools because their family lacked the finances were disadvantaged and described it as a “harsh reality”. This view is supported by previous research, which has shown that social class, housing, and access to resources have considerable impact on academic achievement (Crnic and Lamberty, 1994; Ichado, 1998) and that students who come from families with higher SES are more likely to enrol in elite universities compared to their lower income peers (Li, 2007).

However, not all teachers shared this view. One teacher argued that it was hard to generalise because some students are very committed and are able to make excellent progress in spite of their financial background. She argued that some students were faster learners and that achievement was therefore more dependent on individual ability although she acknowledged that financial background could be an important factor. Another teacher specifically linked financial advantage to the opportunity to attend a private school, where learning English is a requirement.

These results highlight the complex nature of the investigation, in particular because teachers also commented on their views on students, on the curriculum of private and public schools and how socio-economic background may advantage or disadvantage students.

#### 6.4.2. Extrinsic and intrinsic motivational factors

Students’ lack of intrinsic motivation for studying English frequently emerged from the interview. This could be due, in part, to students’ previous learning experience in school and the view of English as a school subject rather than a spoken language. Al-Seghayer (2014, p.19) found that Saudi students often ‘perceive English as a dry and boring subject learned for instrumental purposes, chiefly to pass an examination’.

However, Long (1981) suggests that a foreign language is more than a group of grammar rules to be studied; it is also a dynamic process of communicative competences that allows learners to use language practically, as a tool for communication. Additionally, it is essential to understand that the main aim of learning grammar rules is to employ those rules communicatively in every-day life (Long, 1981). In Saudi Arabia, students generally do not get enough English input outside the classroom nor do they have opportunities for interaction with native English speakers. This can affect their motivation towards studying English at

university and their view of English as a subject rather than a language used for communication in the real world. One teacher claimed that students need self-motivation in order to learn the language, but since most of them do not use the language at home or outside, they lack motivation. Teachers, however, could advise students on the different ways they can use the language outside the classroom. In interviews, it was suggested that students who watched English or American films tended to have a better grasp of the language; that also applied to those who listened to English music. This has also been found in other studies; Kuppens (2010), for instance, found that young people can also acquire language skills by watching foreign media. Exposure to the language, and having the opportunity to employ it, can increase students' intrinsic motivation.

Eccles and Wigfield (2002) argued that 'it is difficult, if not impossible, to understand students' motivation without understanding the contexts they are experiencing' (p.128). For Saudi students, the context does not really present opportunities to use the language in their society. Consequently, Saudi EFL learners lack the internal force and desire for learning the language. Research shows that the classroom environment and school context greatly influence motivation and achievement in complex ways (ibid.).

One of the issues affecting students' motivation, in this particular context, is related to students' unwillingness to speak English. Khan (2015, p. 71), in her study on female students studying in the ELI at KSA, also found 'a strong relationship between the Ideal L2 selves of the students and their achievement in the formal exams', emphasizing the important role of motivation. Previous studies also found that students dislike speaking English because they fear making mistakes in front of their peers. Saudi classmates can be particularly judgemental of each other. Al-Harthi (2016) suggested that the Saudi educational system promotes an ideal view of Muslim women, which is drawn from the aim of portraying women as 'ideal' and 'pure'. For female students in particular, their perceptions of themselves as 'ideal' learners were 'a de-motivating factor in learning and practicing English in the Saudi context' (ibid., p.393). A lack of social encouragement, for instance, caused by a lack of role models or conflict between one's current social identity and one's possible self, have been previously associated with poor ideal second language images (Oyserman and Fryberg, 2006).

One of the teachers interviewed argued that the way the preparatory courses are taught at present does not make it possible to teach the English language courses for specific purposes. Students majoring in the arts and sciences are combined into the same course and therefore, it is not possible to teach the course for specific purposes. She added, in the past, in the college of medicine, teachers used reading texts that revolved around that field. She felt this made the course more relevant to students and it increased their motivation. With regards to teaching English for specific purposes, Jose (2002) argued that ESP learners are better motivated when they see the relevance of what they study to what they want to study. In such cases, the syllabus is usually designed on the basis of their needs. Aside from making the learning experience more relevant and enjoyable, this may also have implications for students' extrinsic motivation. Teachers interviewed commented that they felt Science students were generally more motivated compared to Arts students, because they required a solid knowledge of English for their subject area. Improving their competence in English was therefore perceived as useful and necessary for their field of study. In this respect, one of the teachers commented that Science students work harder, worry about their grades and do well in the exams.

#### **6.4.3. Student preparation**

The interviews emphasised a lack of preparedness for the OOPT, both in terms of familiarity with the test and in terms of subject knowledge and competency in the skills being tested. Teachers also felt that most students were unfamiliar with the content of the OOPT as the topics in the placement tests were not relevant to their previous learning. Students, therefore, took the test without being fully informed about it. It could be argued, however, that, since information on the OOPT is freely available online, students could have utilised resources available to adequately prepare for the test. However, one of the issues remains the significant difference between the content of the OOPT and students' prior learning.

Schönwetter et al. (2002, p. 626) claimed that 'research is limited in understanding the difference between content-familiar and unfamiliar students, moreover the influence that different qualities of teachers have on content-familiar and unfamiliar students is unknown'.

Yet, interviews suggested an overall lack of preparedness for the test. KAU informed all students applying to the university that they would be required to take the OOPT before commencing their studies. The ELI also delivered a presentation, on the day of the examination, to provide students with further information about the administration of the test. Aside from this, however, students did not receive any formal preparation. While the OOPT website *does* provide prospective test takers with some information about the test as well as an opportunity to see sample questions, some teachers felt that this was not sufficient. One of the teachers interviewed commented that the orientation day is very useful because the presentation delivered by the ELI, explains how to navigate through the test. Another teacher also agreed that any sort of orientation would help students become more familiar with the format of the test. She added that she was sure it would have a positive effect, especially because the OOPT is a computerised test and students are used to pen and paper. It was further suggested that the presentation may help to reduce student anxiety about the test because they understand what to expect.

McMillan (2008, p.128) suggested a need to be familiar with the format of the questions to avoid underperformance as a result of poor test-taking techniques. He argued that students should practise the following skills:

- Read or listen to directions carefully
- Read or listen to test items carefully
- Set a pace that will allow adequate time to complete the test
- Bypass difficult items and come back to them later
- Make informed guesses, rather than omitting items.
- Look for grammatical clues to the right answer
- Read all answers before selecting one.

(McMillan, 2008, p.128)

While it may be argued that placement tests do not require specific preparation, students who apply at KAU often have had little or no prior experience with online adaptive tests like the OOPT.

## 6.5. Teachers' recommendations for practice

A number of teachers felt that students did not make adequate progress in the ELI courses and, therefore, those who entered with lower OOPT scores, and progressed to the higher levels by passing the ELI exams, were generally less proficient compared to students who were placed into higher levels (such as levels 3 and 4) as a result of the OOPT. This would seem to indicate that the ELI courses are not effectively developing students' language proficiency as they should. They also may indicate that the ELI assessment may be considerably easier compared to the OOPT.

Mahmoud (2014) explored students' and teachers' views on the foundation year programme at KAU and found that the results of the English language placement test were 'alarming'.

Most of the students lack the basics of English language in the four English language skills. More than 87% of students fall in level one and level two according to the results of the placement test. This is due to the huge gap between the school education and the university education in KSA. The students who were able to reach level three and four were either graduates of private schools or they had the chance to accompany their parents while doing their higher studies in English speaking countries.

(Mahmoud, 2014, p. 337)

The study confirms some of the complaints that teachers raised in the interviews with regards to the weak English language skills student have upon entering the programme. Mahmoud (2014) also found weaknesses in students' proficiency in English, upon completion of the foundation year, suggesting that the ELI modules may not be effective enough to ensure students leave the programme with the necessary skills. Mahmoud (2014), found that staff in other faculties felt that 'students coming from the foundation year lack the minimum of the four English language skills and sub skills that enable them to manage throughout their studies' (Mahmoud, 2014, p.338). It is possible, however, that while students *do* improve their English language skills throughout the ELI courses, because the majority enter university with very weak language skills, they do not have time to make sufficient progress. In his study, some faculty members suggested adding a compulsory intensive summer course in general English and ESP upon completion of the PYP. Mahmoud (2014) thus confirms some of the concerns voiced by teachers with regards to students' progress in the ELI modules. In interviews, one of the teachers commented that students who progressed to level 3 actually have not achieved level 3. She argued that if they took the OOPT again they would

not be placed in level 3 but may actually still be placed in level 1.

Thus, with regards to the effectiveness of teaching in the ELI, a number of teachers argued that students appear to make slow progress throughout the courses and that if they took the OOPT again after completing the course they would probably not score very differently. Al-Bargi (2013) however, seems to suggest otherwise. His study explored students' performance on the listening section of the OOPT within the ELI at KAU. Students took the OOPT twice, once before commencing the ELI beginners course (Level 1) and another time upon completion of the course. He found that 'the increased listening input of the short six-week module appeared to help significantly the learners improve their listening ability and, therefore, their overall level of English use' (p.3582). While Al-Bargi (2013) only explored students' progress in listening (which led to an increase in the overall mark), his findings indicate that further research is needed to explore students' actual progress in these modules. It is possible that teachers interviewed in this study underestimated the actual progress that students make in these short courses, or there may also be differences in practice, motivation and test performance between the male and female sections of the ELI.

The difference in the findings of Al-Bargi (2013) and Mahmoud (2014) appear to suggest a mismatch between teachers' beliefs about students' progress and students' actual progress, which can be measured by looking at test scores. It would be interesting to explore whether test data would present a different picture of students' actual progress in these short courses.

A further concern, which emerged from the interviews, regarded the contact time allocated to PYPs. Some teachers argued that the ELI courses were too short to adequately support students' development. In this regard, one of the teachers commented that ELI courses did not allow sufficient time to make adequate progress. She further added that it is unrealistic to expect that a student can progress from level 1 to level 2 in only 7 weeks. It should be noted, however, that students have 18 contact hours a week in this 12-week period, so this is a very intense course. Previous studies have shown that intensive language courses can achieve comparable results to study abroad (SA) programmes, where students, in addition to being enrolled in a language course, have numerous opportunities to practise the TL outside the classroom. Freed et al. (2004) found that the students in an immersion context (seven weeks of French instruction during the summer, approximately 17.5 hours a week) improved

their fluency more than their peers abroad. Serrano et al. (2011) also found that ‘students’ written and oral performance after an intensive course at home and after the equivalent time abroad was similar’ (p.133). Serrano et al. (2011) concluded that the findings from their study and previous studies on SA appeared to indicate that both intensive classroom practice (as promoted in an intensive course) and real communicative practice outside the class (as encouraged in SA contexts) provide a more suitable environment for L2 compared to ‘regular, ‘drip-feed’ (or not concentrated) L2 instruction’ (p.141). Similar findings also emerged from an earlier study; Carroll (1967) for instance suggests that more time learning the TL usually translates into higher proficiency levels (all other things being equal). Other studies further confirmed the positive effect of intensive exposure to the L2 in classroom settings compared to the ‘drip-feed’ approach characteristic of many EFL programs (Spada and Lightbown, 1989). More recently, White and Turner (2005) found that L1 French-speaking children receiving intensive English instruction in Canada (approximately 400 hours in five months) significantly outperformed their peers receiving regular instruction (36 hours in 10 months).

The number of contact hours students receive in ELI, is comparable to that of short intensive language course and hence, while teachers repeatedly pointed out that they felt 7 weeks were not enough, the lack of students’ progress raises concerns about the effectiveness of the course and possibly other factors, which may affect students’ learning in the ELI courses.

## **6.6. Areas for improvement to support students’ achievement**

The main areas for improvement, which emerged from the analysis, were: increasing teacher consistency across the ELI modules, addressing the strong reliance on textbooks and the possibility of employing an in-house placement test in place of the OOPT. It was also suggested that teachers would benefit from training on the different purposes and types of assessments, so that they could better understand their function within the curriculum.

### **6.6.1. Teacher consistency across ELI modules**

The first area for improvement, which emerged from the interviews, concerns teacher consistency. Some teachers suggested that it would be better to have the same teacher and partner-teacher for the duration of a semester-long course rather than having a different teacher for each short module (as is currently the case). Some teachers argued that having the same teacher would make it easier for students as there would be continuity in the teaching method. One of the teachers felt that having frequent changes in teachers was not helpful. She commented that when you finally get to know your students, they start to talk to you and are encouraged to speak, they are then moved to another course and a different teacher will teach them. She added that it usually takes two weeks for students to get used to their new teacher. Another teacher also agreed that having frequent staff changeovers was not helpful. She argued that students get comfortable with the teacher and the group of students and then they are moved to another level with a different teacher and different students and that is taking her five steps backwards.

Teacher-student relationships are widely recognised as being important to student motivation (Birch and Ladd, 1996; Davis, 2003). Teacher-student relationships have also been previously associated with achievement (Muller, Katz and Dance, 1999; Nieto, 1996). Furthermore, the importance teachers place on developing positive personal relationships with their students has been suggested as one aspect of effectiveness and expertise in teaching (Carr, 2005; Cothran and Ennis, 1997; Davis, 2003; Smith and Strahan, 2004). Addressing the issue of teacher-consistency may be something the ELI could consider; however, it might not always be practical for logistical reasons. It is possible, however, that teachers suggested consistency because it might be more convenient for them. It takes time to get to know students and become familiar with their strengths and weaknesses, therefore, from a teacher's perspective, it would be easier to continue teaching the same cohort of students.



### **6.6.2. Addressing the strong reliance on textbooks in ELI modules**

The second area for improvement relates to teachers' strong reliance on textbooks for the teaching of the ELI modules. The view that ELI modules did not offer sufficient contact time for teachers to cover the curriculum, emerged in conjunction with comments about the textbook, i.e. teachers felt they did not have enough time to cover the whole textbook associated with the level. One of the teachers commented that with the modular system losing a day has a significant impact on the whole pacing of the curriculum. She further added that there is simply not enough time to teach the material because two books are just too much, you can teach the material, but you do not feel students can master it.

Another teacher also shared similar views and felt that the modular system did not effectively support students' learning. She suggested two alternatives: either to adopt a semester system or to keep the current quarterly system but just teach half a book. Both teachers appeared to equate the curriculum with the textbooks associated with the courses. It appeared that teachers interviewed were not used to being able to adapt from the textbook and use the book as a resource rather than trying to teach every page. This suggests a need for adequate teacher training, which would provide teachers with the skills necessary to produce their own teaching resources and rely less on the textbook.

Ironically, the ELI Faculty Handbook 2017-18 ([www.kau.edu.sa](http://www.kau.edu.sa)) specifies the SLOs and states that 'each ELI level has an overarching course goal supported by course objectives, which are supported by the level's SLOs' (p. 18). The individual module syllabi describe the goals and objectives. The course goals relate to the CEFR level, for instance, for level 104, the syllabus makes reference to 'helping learners to achieve an overall English language proficiency leading to higher Independent User of language defined as high B1 level' (ELI 104 Syllabus). This is followed by the Course objectives, which outline what the course is intended to accomplish by the end of the academic module. For level 4 these include: 1. Reading and understanding a wide variety of extended texts, 2. Listening to, understanding and participating in extended oral communications. 3. Constructing a range of coherent and cohesive texts with multiple paragraphs in a fully developed response and 4. Demonstrate consistent control of a wide range of grammatical structures.

There are hence course objectives and course goals guiding these modules, which are not drawn from the textbooks, rather they are informed by the CEFR. What is difficult to determine at this stage, however, is the extent to which these objectives guide teaching practice. These objectives could be met using a variety of different resources but the over-reliance on one textbook seems to suggest that teachers feel they need to cover the whole book in order to adequately prepare students.

### **6.6.3. Employing an in-house placement test**

The third area for improvement emphasised in interviews concerned the possibility of designing an in-house placement test. One of the criticisms voiced by teachers concerned the OOPT's relevance. Some teachers argued that the content of the OOPT was not relevant to students' prior learning or to what is taught in the ELI modules and thus suggested designing an in-house placement test.

According to (Hughes, 2003, p. 5), teachers can make three important contributions to the improvement of testing: they can write tests themselves; they can inform other people who are involved in testing processes; and they can put pressure on examining boards and professional testers to improve their tests. With reference to placement tests specifically, Hughes (2003, p.17), argues that the most successful ones are those designed for particular situations. They rely on the key features at different levels of teaching in the institution and are therefore 'tailor-made'. This usually means that they have been produced 'in house' and not designed by a publisher. While it is time-consuming to design a placement test that specifically addresses the features of each level, it can save time and effort later on, through accurate placement (*ibid.*). Teachers interviewed had mixed views on whether an in-house placement test might be an improvement over the OOPT.

However, if in-house placement tests are to be used in place of the OOPT, it is essential that they are very carefully designed. Experienced and creative teachers, who have an interest in writing tests, may be the best test designers, provided they receive the necessary training.

Unlike test developers, teachers are well aware of students' strengths and weaknesses as well as students' interests at a particular level. Regrettably, there is a shortage of assessment experts in Saudi Arabia (Rauf, 2015), which affects the choices that universities have when it comes to placement tests.

If the ELI wishes to continue using the OOPT, teachers need to become more familiar with the content. Shimizu (2002), in a study carried out at a Japanese university, found that one of the main advantages of having placements tests was that it allowed teachers to prepare appropriate materials for the level they were about to teach. Where teachers have no knowledge of the placement test's content, the extent to which it can inform practice is limited. Becoming familiar with the placement test would thus allow teachers to fully understand what skills their students need to develop and how the university can ensure students are fully informed of the OOPT before they take the placement test.

## **6.7. Summary**

In interviews, teachers described their perceptions of the OOPT as a placement test, such as its perceived effectiveness and concerns about its cultural appropriateness for Saudi students, as well as commenting on test use and perceived consequences such as misplacement. According to Messick (1989) these findings can provide information about the OOPT's validity.

The relationship between placement tests and the context in which these are implemented constituted a key finding. The cultural context and education system in Saudi Arabia, also emerged of significant importance as did students' learning experience prior to university, which teachers felt influenced attitudes towards learning English and achievement in the placement test. Furthermore, as a result of the learning context, teachers felt that some students had very weak computer literacy, which also played a role in achievement. However, the context of the study was found to influence more than students' achievement. Teachers' views of the placement test, for instance the criticism of the cultural content and the view that the OOPT was not perceived to be culturally relevant to Saudi students, were similarly shaped by the cultural context in which they are teaching. Surprisingly, non-Saudi teachers

generally agreed with some of the beliefs of Saudi teachers, for instance, that Science students are better than Arts students. The findings also suggest the need for in-service work with teachers in order to familiarise them with the various purposes of assessment. The apparent lack of familiarity with assessment types and purposes could be addressed through adequate teacher training. A better understanding of the OOPT and the other assessments taking place in the ELI, could help teachers reflect on their students' weaknesses and how best to support their learning.

## **CHAPTER 7 CONCLUSION AND RECOMMENDATIONS**

### **7.1. Introduction**

The chapter opens with a reflection on the extent to which the findings provided evidence in support of the research question. It then illustrates the study's contributions to knowledge in the area of English language placement tests and, more specifically, on teachers' perceptions of the test, its use and consequences.

Furthermore, the study highlights the important role of teachers' beliefs about the test and about students in the PYP at KAU. The chapter further discusses recommendations directly related to the research questions as well as wider implication for teaching and learning. It concludes with a discussion on the suggestions for future research and limitations of the study. This study set out to explore teachers' view of the OOPT and factors, which, according to teachers, affected students' achievement on the placement test and, more generally, their performance in ELI modules.

### **7.2. Revisiting the research questions**

The investigation aimed to explore ELI teachers' views on the OOPT as a placement test, on factors affecting achievement, the ELI modules and ways in which the system can be improved to better support students.

While the original aim of the study was to focus on a narrow aspect of the validity of the OOPT, as previously mentioned in chapter 4, as a result of the type of data collected and the challenges of collecting data from students, there was a shift in focus towards a more exploratory type of investigation of validity. The findings, provide insight into teachers' perceptions, uses and consequences of the OOPT, which provide a perspective on the test's validity in line with Messick's concept of unified validity.

Additional insight emerged from the data, which did not specifically address the research questions, but rather illustrated some of the complexities of the research setting. For instance, the findings exposed a possible lack of understanding, on behalf of some of the teachers, with regards to the differences between placement tests and other types of tests, such as achievement tests or pre-arrival tests, which aim to assess a students' proficiency in the language (such as the IELTS or TOEFL). This is a key finding, as it affects the way teachers' views on the OOPT may be interpreted. Teachers' opinions on the OOPT are shaped by their knowledge of it and, therefore, where teachers made criticisms of the test, this needs to be taken into perspective. Teachers may have expected the OOPT to test different skills or include a different type of content because they did not fully understand the purpose for which the OOPT was designed.

These findings brought me to reflect upon how my own views on placement tests evolved throughout my doctoral journey. Before undertaking this investigation, my own view of the kinds of skills that placement tests should include did not significantly differ from those of my colleagues within the ELI. I was also of the opinion that the placement test should test all four skills, however, after undertaking this research, it became clear that there are different types and purposes of placement tests. I now recognise that one should first understand the purpose for which a test was designed before making conclusions about which skills it should test or whether relevance to the course content is indeed necessary.

#### **7.2.1. Research question 1: How do teachers view the OOPT?**

Teachers held mixed views on the OOPT; some teachers described the placement as effective, while others felt that it was not culturally relevant to Saudi students or that students' lack of familiarity with online testing could have disadvantaged them, thus affecting the reliability of the scores.

The findings also exposed the extent to which teachers were unfamiliar with the OOPT. Teachers who were familiar with the test generally agreed that the OOPT was an **effective tool** and **placed students correctly**, while those who were unfamiliar were neutral in their

responses. Similar results were found when teachers were asked about whether the OOPT **reflects students' language ability**. The familiarity group agreed on this point, while the unfamiliarity group remained neutral.

With reference to questions relating to the skills tested in the OOPT, teachers held mixed views on whether the OOPT should test the productive skills of **speaking and writing** in addition to reading and listening, which are both receptive skills. Some teachers felt that the OOPT should test all **four skills** in order to provide a **fair** measurement of students' language ability.

Some teachers commented that some students were **misplaced** as a result of taking the OOPT and argued that in some cases students may have underperformed in the placement test because of **weak computer literacy** or unfamiliarity with the **cultural content**.

### **7.2.2. Research question 2: How does the OOPT impact on placement and teaching, in the opinions of teachers?**

Generally, teachers who were familiar with the OOPT held positive views on the test and felt that students were **being placed in a level suitable for their ability**. Some of the teachers interviewed disagreed and argued that in some cases students underperformed (false negative) or they were placed in a level too high for their ability (false positive). Criticism of the OOPT also concerned the **cultural content** and the fact that students needed good **computer literacy** skills in order to be able to perform to the best of their ability. Some felt that adequate **preparation for the test** would help ensure students perform better in the OOPT, which could improve the extent to which student are accurately placed. In interviews, some teachers complained about the degree to which students within one module had very different levels of competence in English. Some attributed this to the OOPT and felt that students were being misplaced as a result of the placement test.

The way in which the ELI employs the OOPT results to place students into the four **ELI modules** also emerged as problematic, since the OOPT recommends placing students into six different modules. Because the ELI uses the marks differently from the recommendations of the test developers, it is possible that some students may be misplaced and that higher

levels may include students of very different ability. However, because the majority of students are placed in either level 1 or 2, it is understandable that the ELI decided not to offer any level higher than level 4.

### **7.2.3. Research question 3: What are the factors that might affect students' achievement in both the OOPT and during the English module, in the opinion of teachers?**

Several teachers felt that **Science students** generally do better compared to **Arts students**. Similarly, many agreed that those who had attended **private schools** generally held more positive attitudes towards studying English and were better prepared. A further novel finding revealed criticisms of the **modular system**. Some teachers felt that the main issue was that the courses were too short for students to make adequate progress. They argued that this led to a mismatch between students who progressed to higher levels by passing the ELI assessments and those who were placed in higher levels as a result of having achieved a higher score in the OOPT. Both the questionnaire and interview results emphasised the importance of student **motivation**. The majority of survey respondents agreed that motivation plays a vital role in students' success. In interviews, some teachers suggested that students' lack motivation in the ELI modules could be due to the fact that ELI modules are not taught for specific purposes and hence the content is not subject specific. Other perceived factors affecting students' achievement included **stress and anxiety, attitude towards language, language background**, as well as **socio-economic** factors and **health**. Teachers felt that students who had started learning English from an **earlier age** were generally advantaged. Surprisingly, there was a difference between the questionnaire and interview responses on the topic of financial background. While the majority of questionnaire respondents did not perceive that financial background was strongly linked to improved performance in the OOPT and ELI modules, interview respondents appeared to equate higher financial status with students' opportunity to attend a private school. The interview respondents thus agreed that students from families with higher SES generally did better in the OOPT and had better English language skills upon entry.



#### **7.2.4. Research question 4: How can the system be improved to support students' achievement?**

Some teachers suggested adopting a **semester system** rather than a modular system as this would allow teachers more time to adequately prepare students for the next level. Some argued that this would also allow the same teacher to continue working with the same group of students, which would be preferable to the current frequent staff changeovers. Teachers also argued that despite using a placement test, there are differences between the competences of students, and this poses challenges for teaching. This finding seems to suggest that teachers might be unfamiliar with differentiation techniques that could be employed when teaching **mixed-ability** classes. Teachers suggested employing an alternative test, such as an **in-house placement test**, which would be designed to match the Cambridge special edition books used to teach the ELI modules. While these were recommendations made by teachers in interviews, there are other recommendations, which emerged from the data. These are discussed later in the chapter.

### **7.3. Other emerging findings**

1. A lack of understanding on the differences between placement tests and other tests, such as proficiency or achievement tests emerged from the interviews. Teachers' assessment literacy also emerged as generally weak.
2. Over-reliance on textbooks also emerged from the interview data as a key finding. With reference to the ELI modules, teachers commented that they did not have sufficient time to 'cover the whole book'. This seems to suggest that teachers view the textbook as the curriculum rather than looking at the course objectives specified in the module syllabus, to determine what their students should achieve by the end of the module.

Reference to these findings is further made in the chapter with regards to the recommendations.

#### **7.4. Contribution to knowledge**

The study addresses a gap in the literature by evaluating the perceived effectiveness of employing the OOPT as a placement test in Saudi Arabia. It further contributes to the current literature on PYPs and exposes teachers' views on factors affecting students' performance in placement tests as well as on the English language modules. The research additionally provides an overview of some of the issues that teachers experienced as a result of teaching students who had been placed in certain English language modules, on the basis of their OOPT scores. More generally, the study adds to our understanding of teachers' beliefs about placement tests and factors affecting students' achievement.

Unfortunately, few studies have been carried out on PYPs in Saudi Arabia and to my knowledge this is the only study, which has specifically explored teacher's views on the OOPT in Saudi Arabia. The literature on HE teachers' understanding of assessments practices is also scarce and 'little is known about the ways in which they understand the different conceptual models that can be found in the assessment literature' (Sadler and Reimann, 2018, p. 134).

The context of the study played a significant role in the investigation and hence there are contributions to knowledge, which are specifically relevant to the Saudi context. Teachers' beliefs, regarding the use of placement tests and the relationship between the characteristics of students and their achievement, provide valuable insight for the development and improvement of PYPs in Saudi Arabia. The present study makes several noteworthy contributions, for instance, it exposes ways in which Science and Arts specialisations are viewed both in secondary education and tertiary education in Saudi Arabia. Teachers' beliefs concerning the preparedness of students coming from private versus public secondary schools is also a finding, which is relevant within this context. The extent to which the curriculum in private schools differs from that of public schools is an important aspect of the Saudi education system. Some of the study's contributions to knowledge, therefore, are specifically relevant to Saudi tertiary education and can inform current practice within this context.

There are other aspects of the study, however, which are not exclusively relevant to the Saudi context. For instance, the study illustrates the importance of understanding the different types and purposes of placement tests. Strikingly, the findings highlight that teachers who were more familiar with the OOPT generally rated it better. This has implications for teachers' professional development and, more specifically, for the need to familiarise themselves with placement tests.

### **7.5. Contribution to theory**

An alternative view on the way placement tests are currently perceived emerged from a careful analysis of the study's findings. Teachers' views on using placement tests in real-world scenarios add to our understanding of the complexities of administering such tests in certain cultural contexts. While the literature differentiates between placement tests with a proficiency orientation and those, which reflect the content of language courses, it is argued that this dichotomy is not necessary.

Wall et al. (1994) distinguish between two different types of placement tests, namely:

- Placement tests with a proficiency orientation (such as the OOPT)
- Placement tests developed to specifically reflect the nature of language courses

They argue that the content of the former 'bears no direct relationship to the content of the language courses on to which students are placed' (p.322), while the latter 'is intended to reflect the nature of such language courses and might be said to be 'preachievement'' (ibid.).

This dichotomous view of placement tests is not helpful; it may in fact be better to design a placement test with a proficiency orientation, but which also reflects, at least to some extent, the nature of language courses. In interviews, teachers criticised the perceived lack of relevance between the OOPT and the ELI curriculum and suggested adopting a placement test, which would be more relevantly linked to the ELI modules. Since every language test needs to have a content of some kind, it would be preferable to align this content of the placement test to the curriculum. This dichotomous view of placement tests being *either*

proficiency oriented *or* reflecting the nature of language courses is not beneficial to test developers, who may be interested in exploring ways in which placement tests can be better matched to course curricula. With reference to the Saudi context, we see that there is a demand for English language materials, which are designed to be more culturally relevant, such as the special edition textbooks used in the ELI. At the same time, teachers complained about students being misplaced, thus highlighting the importance of placement tests with a proficiency orientation. Recommendations for test-developers are further discussed later in the chapter.

There are however also challenges in employing materials, which have been specifically designed for the Saudi context. As previously mentioned in chapter 6, Sulaimani and Elyas (2018) found that special edition textbooks, such as the *New Headway*, misrepresent women in Saudi Arabia. The authors concluded that the publishers should ‘glocalize’ the *New Headway Plus: Special Edition* series by ‘incorporating new educational ideologies’ characteristic of a more modern Saudi Arabia (p.72). The term ‘glocalize’ refers to finding a middle ground between the concept of ‘global’ and ‘local’ (Selvi and Rudolph, 2018). Hence, while adapting materials to meet the needs of the learners may appear advantageous, a number of difficulties arise when attempting to portray an authentic representation of a multicultural Western society in a culturally sensitive way.

## **7.6. Recommendations and implications**

This section firstly describes recommendations directly related to the research questions, and secondly those which emerged from the interview data. This is followed by a discussion on the wider implications for teaching and learning.

### **7.6.1. Recommendations drawn from findings related to the research questions**

The following recommendations are drawn from findings directly related to the research questions. Recommendations include training needs for teachers as well as considerations about in-house placement test and alternatives to the OOPT.

1. When teachers were asked about how they viewed the OOPT, it emerged that many of them were **not familiar** with the test and could therefore only provide limited comments about their knowledge or experience of it. It is therefore necessary that teachers receive specific training on the OOPT so that they are fully aware of the scope for its use within the ELI and the ELI's rationale for choosing the OOPT over other placement tests. The training could take place in the form of professional development courses or workshops, which could provide teachers with the opportunity to look at the content of the OOPT and share views on possible links with what they teach in the ELI modules.
2. Some teachers felt that the ELI modules could support students' achievement better, if a **semester system** was instructed instead of the modular system. The ELI course leaders should consider whether a semester system could be introduced, as well as the possibility of increasing the contact hours for each level, so that teachers would have more time to support students' learning.
3. Some teachers felt that in some instances students were misplaced as a result of the OOPT and suggested adopting an **in-house placement test** in its place. Therefore, some teachers argued that an in-house placement test might be more effective, as this could be structured to better match the curriculum as well as students' skills and knowledge. There are, however, both advantages and drawbacks associated with designing an in-house placement test. Teachers could be involved in the process of designing the placement test and gain knowledge by doing so. This might be an effective way of developing teachers' knowledge on the different types and purposes of placement tests. It has been previously pointed out, that teachers' lack of involvement can result in frustration and demotivation, which can result in loss of job satisfaction (Collie et al., 2012). A further advantage of in-house placement tests is that they can be developed to 'reflect course content more loosely' (Green, 2012, p.168). Designing a placement test, however, is a time-consuming job and there is the risk that mistakes could be made. To consider the option on an in-house placement test, the ELI should first evaluate whether staff have sufficient expertise in

test design. Other alternatives might include adopting the CEPT or experimenting with both the CEPT and the OOPT to evaluate the differences between them. It is possible that the CEPT might be perceived as more relevantly linked to the current curriculum since it is published by Cambridge.

#### 7.6.2. Recommendations not drawn from the research questions

1. Despite the advantages of placement tests such as the OOPT, which have been professionally designed by test developers, there still remain some concerns with its applicability to the research context. A key concern is the fact that **teachers in the ELI are not actively involved in the selection**, administration or evaluation of the OOPT. While it is understandable that the decision regarding which placement test the university should adopt would lie with the ELI directors, teachers' lack of familiarity with the OOPT resulted in a misunderstanding of what the OOPT was intended to measure. It is thus recommended that teachers **play a more active role in decisions concerning which placement tests** to adopt and the rationales behind these decisions.
2. Some teachers viewed the OOPT as an achievement test rather than a placement test and did not appear to fully understand the **distinction between the two types of tests**, thus highlighting a weakness in assessment literacy.  
Stiggins (1995) argued that in order to help students attain higher levels of academic achievement, teachers need to develop appropriate types and levels of assessment literacy.
3. The findings also exposed an **over-reliance on the textbook**; in fact, a number of teachers expressed concerns about not having enough time to teach the whole book. While the course syllabi for the ELI modules set out course objectives, which are independent from the textbook, it is argued that teachers may require additional training to explore alternative ways to ensure the course objectives are met. It is recommended that ELI directors organise meetings prior to the commencement of the

academic year in order to discuss the learning outcomes specified in the module syllabus. These training sessions should create an opportunity for teachers to discuss how to structure teaching around the learning objectives and make creative use of alternative materials to reduce the reliance on textbooks.

4. Teachers could also observe **mixed-ability lessons** in order to reflect upon how they could use differentiation in their teaching. The ELI could also offer workshops led by more experienced teachers on mixed-ability groups and differentiation strategies. It might also be helpful to create a database of resources that teachers could use, or adapt, in order to support weaker or more advanced learners within the level they teach. Understanding students' OOPT scores might also be helpful to teachers, as it may give them an indication of which students scored higher and lower within the same level.
5. The apparent mismatch in students' ability in the ELI modules seems to indicate that it is **easier for students to pass the ELI module** assessments compared to obtaining a high score in the OOPT. This issue could be explored further, as it is difficult to establish whether this may be linked to teachers' perception of students' learning or whether teaching itself is not effective enough to sufficiently support students' progress. As mentioned in chapter 7, Al Bargi (2013) found that ELI students in Level 1 made significant progress in their listening skills within the short six-week module, however Mahmood (2014), who, in similar vein to this investigation, explored the views of teaching staff across other faculties, found that the majority felt that students had not developed the necessary English language skills upon completion of the PYP. The difference in the findings of these two studies seems to suggest a mismatch between teachers' intuitive judgments and impressions and hard evidence. Perhaps it would be necessary to measure students' actual progress and make the data visible to both teachers and students, so they can note progress or focus on what needs improvement.

6. The ELI could re-evaluate the current use of the **OOPT thresholds** used to determine the score range for each corresponding CEFR level and ELI course. As previously mentioned, the OOPT has its own recommended CEFR levels but the ELI employs the scores differently. In interviews, some teachers argued that there was a wide range of ability within the courses they teach and therefore it might be worth exploring whether six different levels, rather than four, might address some of these issues and be able to better support students' learning. If new courses following the OOPT recommended thresholds are offered, these could be piloted for a year and then assessed to evaluate their effectiveness. However, the decision to offer two additional ELI modules would need to be justified. For instance, the ELI would need to have a sufficient number of students who could be placed in the two higher levels. However, regardless of how accurately the placement test scores are used to place students into an appropriate language course, the risk of misplacement would still be present. Furthermore, there would still be differences in ability among students, since some would have scored in the lower-range, while others would have scored higher. Revisiting the way in which scores are translated into CEFR levels within the ELI may improve current practice, but teachers will still be faced with mixed-ability groups.

### 7.6.3. Wider implications for teaching and learning

1. Test developers should consider the possibility of designing a **special edition placement test**, which would match the curriculum of the textbook. Publishers could design a special edition of the OOPT to match the content of the Oxford special edition materials as well as a special edition of the CEPT to match the Cambridge special edition materials. Institutions could then evaluate which placement test and its associated teaching materials are best suited for their English language courses. Such a placement test would still be guided by a professional orientation but would also be designed to reflect the nature of the language course.
2. An important implication for practice beyond the ELI context concerns the **need for language teachers to become more familiar with the differences between**



**placement** tests and tests developed for other purposes. While the scores of proficiency tests, such as the IELTS and TOEFL, have at times been used for placement into language courses, these have not been designed for this purpose by their developers. Language teachers would benefit from specific training on the different types and purposes of placement tests in order to develop a better understanding of what is being tested and why. Crusan (2014) highlights that even language program directors ‘often struggle with placement procedures’ and therefore are required to ‘frequently revisit their decisions about many aspects of placement’ (p.5). Thus, it is understandable that most teachers would not be familiar with the differences between the purposes and characteristics of placement tests and those of similar tests, such as proficiency tests.

3. The importance of **teachers’ beliefs about test performance**, placement tests and about students more generally needs to be better acknowledged. The study emphasised the important role of teachers’ beliefs and the extent to which teachers’ beliefs are shaped by previous learning experience and teaching practice. Language programme directors could gain important insight into some of the challenges of placement tests by listening to the experiences of teachers.
4. A further implication concerns exploring ways to **increase relevance** between the **placement test** and the **curriculum** taught in the ELI modules. In the 2015-2016 academic year, the ELI decided to implement a change in curriculum, which resulted in replacing the previously used New Headway Special Edition textbooks, published by Oxford, with the Cambridge Special Edition ones. Some teachers felt that this change was not beneficial as there was greater consistency between the placement test and the curriculum when they were using Oxford materials. While the OOPT is exclusively employed to determine in which course students should be placed, for pedagogical reasons, employing textbooks from the same publisher might increase relevance. Creating an interface between the curriculum and the placement test would increase the relevance between what is initially tested and what is later taught in language courses.

## 7.7. Suggestions for future research

The study also identifies key recommendation for future research both in the field of placement tests and teachers' beliefs about student performance. The main recommendations for research are summarised as follows:

1. As the review of the literature on cultural appropriateness and adaptation of placement tests has shown, there are no 'off the peg' solutions that will address this issue quickly and easily. However, it is important for placement test designers to take account of these issues rather than ignoring them. The findings suggest a need for a new but modern adaptation of the presentation of culture in placement tests, which, as described in chapter 3, should reflect a glocal perspective. Test designers should also consider the risks of teachers and students 'revering' presentations of culture in tests and consider including test items which foster a more critical perspective on the target culture(s).
2. There is a need to gain a better understanding of the challenges experienced by students when taking a placement test such as the OOPT. While this study presents the perspective of teachers, further research is needed to explore how students themselves feel about the OOPT and the ELI modules. As with teachers' beliefs, learners' beliefs are similarly shaped by a variety of factors, including 'the learner's previous learning experience' (Sawir, 2005, p. 571). The type of teaching that learners experience in secondary school can shape their beliefs about teaching as they enter HE, thus, as Mori (1999) suggests, 'language teachers need to remember that what they do in the language classroom continues to shape students' beliefs and expectations about learning' (cited in Sawir, 2005, p. 571). Exploring students' experiences and views on the OOPT as well as on the different teaching approaches experienced in school and in the ELI modules would help identify key challenges students face in their first year of university.
3. Further research may also explore whether students in private universities experience similar challenges with the OOPT compared to those in public universities. In Saudi

Arabia, non-Saudi students are only allowed to study in private universities, which means that all students in public universities are Saudi nationals, hence the student population in private universities is likely to be much more diverse.

4. Furthermore, with regards to the methodology, while the study included both qualitative and quantitative data, the interviews were not administrated with the aim of following up on the questionnaire responses. Additionally, the sample size, in particular for the questionnaire, was small, which had implications for the extent to which the data could be statistically analysed as well as for the conclusions that could be drawn from it. Future research could use a similar methodology but with follow-up interviews, which could explore the questionnaire responses further. Furthermore, observation of test taking could further illustrate some of the challenges students experience in the process.
5. The study furthermore exposed teachers' views on the importance of preparedness and motivation for successful test performance. Future research could explore the extent to which prospective students in Saudi Arabia wish to perform well in the OOPT or other placement tests used in preparatory year programmes. Additionally, it would be interesting to examine whether gaining a high score in an English language placement test is considered important and what strategies test-takers employ in order to adequately prepare for placement tests.
6. There is also a need to explore the degree to which teachers, who teach English in secondary schools, are informed of the university admission requirements in Saudi Arabia. Considering that, as emerged from this study, the approach towards teaching English in Saudi government high schools significantly differs from the approach adopted in HE, school teachers may also require adequate training in order to fully understand the type of competencies that students need to develop.

## 7.8. Limitations

The small scale and exploratory nature of the present study inevitably poses limitations, which affect the extent to which the findings may be generalised. The key limitations of the investigation can be summarised as follows:

1. As the study was carried out within the ELI of one university, the findings may not represent the perceptions of teachers across other institutions.
2. Furthermore, only female staff members who teach the female students in the ELI were interviewed and it is possible that male staff within the same university do not share the same views or experiences.
3. The findings exclusively portray a teachers' perspective on the OOPT, the ELI modules and factors affecting students' achievement, yet students may hold different or even opposing views on the subject. Therefore, including a student perspective would be beneficial in gaining insight into the challenges that students experience, their learning needs and their own views on this test.

Aside from the limitations mentioned above, it is further important to acknowledge both the strengths and drawbacks of my positionality as a researcher. Hammersley (1993) suggested that the results of research may differ depending on the individual undertaking it, thus highlighting the importance of reflexivity. Differences might be small, but they would still be present. Different researchers may, for instance, place greater emphasis on some matters compared to others. One of the factors, which may have influenced the approach I took towards my research concerns my position as an insider. The advantages and disadvantages of insider research as well as my rationale for adopting an insider's perspective are discussed in detail in chapter 4. Conducting the research from an insider's perspective offered several advantages, such as already being aware that some teachers were critical about the OOPT as well as being familiar with the staff, the context of the university and the curriculum of the ELI. Being familiar with the context, and with some of the informal discussions shared with colleagues regarding the OOPT, prompted me to investigate the issue further. My positionality, however, also had its drawbacks. For instance, all the participants in my study were female teachers. Furthermore, undertaking research 'at home' also brings in different dynamics with regards to 'concerns of insider-outsider and politics of representation'

(Sultana, 2007, p.378). With reference to my study, while I interviewed colleagues within the ELI, some of the teachers interviewed were less experienced or had joined the ELI fairly recently. Sultana (2007) suggests that when conducting insider research 'it is critical to address the researcher/research participants binary, and renegotiate power relations, responsibilities and hierarchy within the constraints and contexts of any given research endeavour' (p.381-382). The possibility that power relations affected the trustworthiness of the responses of some of the participants can be regarded as a limitation of the study and of insider research more generally. For instance, the recommendations described earlier in this chapter are clearly influenced by my familiarity with the ELI. My ability to relate to some of the teachers' comments, and identify with what they said, may also have influenced the extent to which some findings may have been given greater importance over others.

## **7.9. Conclusion of the Thesis**

The findings of the thesis illustrate the interface between placements tests and their use in real world settings. On the one hand, they highlight the tensions behind developing a test applicable to a range of groups, contexts and learners. On the other, they emphasise the multifacetedness of the contexts and the complexity of the ways in which these tests are interpreted, used and how they are linked to teaching practice. A linear model of simply designing a placement test, using it and placing students is not borne out by observation of practice; many factors and issues intervene and make this a lot more complicated.

The findings thus place emphasis on the complexities of the environments in which teachers have to negotiate when placing, teaching and assessing students or designing curricula, as well as the difficulties arising from the situation in which they find themselves and the skills and understandings required to be able to be effective in such environments.

This study can serve as a starting point for future research, as it exposes a number of issues, which are currently under-researched in Saudi Arabia. It is hoped that the recommendations drawn from the findings serve to improve current practice with regards to the selection and administration of English language placement tests as well as raise awareness about the complexities behind the use of placement tests in real world contexts.

## Bibliography

- Abell, S. K., & Siegel, M. A. (2011). Assessment literacy: What science teachers need to know and be able to do. In D. Corrigan, J. Dillon, & R. Gunstone (Eds.), *The professional knowledge base of science teaching*, 12, 205–221.
- Abu-Rabia, S. (1998). Social and cognitive factors influencing the reading comprehension of Arab students learning Hebrew as a second language in Israel. *Journal of Research in Reading*, 21(3), 201–212.
- Ahmad, J. (2015). Traditional & socio-cultural barriers to EFL learning: A case study. *English Language Teaching* 8(12), 191. <http://dx.doi.org/10.5539/elt.v8n12p191>
- Ahmadi, H., & Shah, S. R. (2014). EFL textbooks: Exploring the suitability of textbook contents from EFL teachers' perspective. *VFAST Transactions on Education and Social Sciences*, 5(1), 12-20.
- Al Sadaawi, A. S. (2010). Saudi National Assessment of Educational Progress (SNAEP). *International Journal of Education Policy and Leadership*, 5(11).
- Al-Adwani, S., & Al-Abdulkareem, A. (2013, June 29). English: A nightmare for students of the preparatory year programs. *Al Madina*. Retrieved from <http://www.al-madina.com/node/462857>
- Al-Asmari, A. (2013). Saudi university undergraduates' language learning attitudes: A preparatory year perspective. *International Journal of Asian Social Science*, 3(11), 2288–2306.
- Al-Bargi, A. (2013). Listening in beginner language courses: The gathering storm in the Saudi context. *Life Science Journal*, 10(4), 3580–3584.
- Al-Juhani, S. H. H. (2012). A study of failure and withdraw phenomena at College of Preparatory Year, Princess Norah bint Abdulrahman University. *The Future of Arabic Education*, 19(79), 245–334.
- Al-Liheibi, A. H. N. (2008). *Middle and high school students' attitudes toward physical education in Saudi Arabia* (Unpublished doctoral dissertation). University of Arkansas, Fayetteville, AR.
- Al-Maliki, T. A. (2013, June 8). The preparatory year programs at universities: A superior goal but! *Al-madina*. Retrieved from <http://www.al-madina.com/node/458408>
- Al-Otaibi, M. N. (2015). *A proposal for the management improvement of preparatory year programs at Saudi universities guided by current international experiences*

- (Unpublished doctoral dissertation). Al-Imam Muhammad ibn Saud Islamic University, Riyadh, Saudi Arabia.
- Al-Sadan, A. (2000). Educational assessment in Saudi Arabian schools. *Assessment in Education: Principles, Policy and Practice*, 7(1), 143–155.
- Al-Saraj, T. M. (2014). Foreign language anxiety in female Arabs learning English: Case studies. *Innovation in Language Learning and Teaching*, 8(3), 257–278.
- Al-Seghayer, K. (2011). *English teaching in Saudi Arabia: Status, issues, and challenges*. Riyadh, Saudi Arabia: Hala Printed Co.
- Al-Seghayer, K. (2014). The actuality, inefficiency, and needs of EFL teacher-preparation programs in Saudi Arabia. *International Journal of Applied Linguistics and English Literature*, 3(1), 143–151.
- Al-Shehri, S. (2017). A developmental paradigm for English language instruction at preparatory year. *Arab World English Journal*, 8(3), 432–447. Retrieved from [https://www.researchgate.net/publication/320286129\\_A\\_Developmental\\_Paradigm\\_for\\_English\\_Language\\_Instruction\\_at\\_Preparatory\\_Year\\_Programs](https://www.researchgate.net/publication/320286129_A_Developmental_Paradigm_for_English_Language_Instruction_at_Preparatory_Year_Programs)
- Al-Zarah, L. N. (2008). *An investigative study of the female initial teacher preparation programme in Saudi Arabia in an Islamic context: Perceptions of the key participants* (Unpublished doctoral dissertation). University of Durham, Durham, UK.
- Alamri, M. (2011). Higher education in Saudi Arabia. *Journal of Higher Education Theory and Practice*, 11(4), 88–91. Retrieved from <http://search.proquest.com/docview/923609954?accountid=17227>
- Alaqeeli, A. S. (2014). The preparatory year: Global perspectives and local practices. *The Saudi Journal of Higher Education*, 11, 45–64.
- Alderson, J. C. (2004). Foreword. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.). *Washback in language testing: Research contexts and methods*. London, England: Lawrence Erlbaum.
- Alderson, J. C., & Banerjee, J. (2001). State of the art review: Language testing and assessment (part one). *Language Teaching*, 34(4), 213–236.
- Alderson, J. C., & Banerjee, J. (2002). State of the art review: Language testing and assessment (part two). *Language Teaching*, 35(2), 79–113.
- Alderson, J. C., & Wall, D. (1996). Editorial. *Language Testing*, 13(3), 239–240.

- Alderson, J. C., Clapham, C., & Wall, D. (1995). *Language test construction and evaluation*. Cambridge, United Kingdom: Cambridge University Press.
- Alderson, J.C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14(2), 115–129.
- Alexander, R. (2000). *Culture and pedagogy: International comparisons in primary education*. Oxford, England: Blackwell.
- Alfahadi, A. (2014). *Saudi teachers' views on appropriate cultural models for EFL textbooks: Insights into TESOL teachers' management of global cultural flows and local realities in their teaching worlds* (Unpublished doctoral dissertation). University of Exeter, Exeter, United Kingdom. Retrieved from <https://ore.exeter.ac.uk/repository/bitstream/handle/10036/3875/Abdulrahman%20Alfahadi%20PhD%20thesis%202012%20Final.pdf?sequence=1>
- Alhawsawi, S. (2013). *Investigating student experiences of learning English as a foreign language in a preparatory programme in a Saudi university* (Unpublished doctoral dissertation). University of Sussex, Brighton, UK.
- Alharthi, T. (2015). Reasons for Vocabulary Attrition: Revisiting the State of the Art. *International Education Studies*, 8(10), 86-95. doi: 10.5539/ies.v8n10p86
- Al-Jabri, N. (2003). *Quality and efficiency of Saudi Education: An investigation into boys' secondary schools*. (Unpublished doctoral dissertation). The University of Manchester, UK.
- Aljumah, F. (2011). Developing Saudi EFL students' oral skills: An integrative approach. *English Language Teaching*, 4(3), 84–89.
- Alkharashi, M. A. (2012). The oil economy of Saudi Arabia. *Proceedings of the Academy for Economics and Economic Education*, 15(2), 27–30.
- Almutairi, N. H. (2008). *The influence of educational and sociocultural factors on the learning styles and strategies of female students in Saudi Arabia* (Unpublished doctoral dissertation). University of Leicester, Leicester, UK.
- Alrabai, F. (2014). A model of foreign language anxiety in the Saudi EFL context. *English Language Teaching*, 7(7), 82–101. <http://dx.doi.org/10.5539/elt.v7n7p82>
- Alrabai, F. (2015). The influence of teachers' anxiety-reducing strategies on learners' foreign language anxiety. *Innovation in Language Learning and Teaching*, 9(2), 163–190. DOI: 10.1080/17501229.2014.890203



- Alrabai, F. (2016). Factors underlying low achievement of Saudi EFL learners. *International Journal of English Linguistics*, 6(3), 21–37.
- Alrahaili, M. (2013). *Predictors of L2 attitudes and motivational intensity: A cross-sectional study in the Saudi EFL context* (Unpublished doctoral thesis). The University of Newcastle, Callaghan, Australia.
- Alrashidi, O., & Phan, H. (2015). Education context and English teaching and learning in the Kingdom of Saudi Arabia: An overview. *English Language Teaching*, 8(5), 33–44.
- Alshahrani, M., & Alandal, A. (2015). An investigation of anxiety among elementary school students towards foreign language learning. *Studies in Literature and Language*, 11(3), 29–40.
- Alt, M. N., & Peter, K. (2003). Private schools: A brief portrait. In *The Condition of Education 2002*. Washington, DC: US Department of Education.
- Alvesson, M. (2003). Beyond neopositivists, romantics, and localists: A reflective approach to interviews in organizational research. *Academy of Management Review*, 28(1), 13–33.
- American Educational Research Association., American Psychological Association., National Council on Measurement in Education. Joint Committee on Standards for Educational and Psychological Testing (U.S.) (1999). *Standards for educational and psychological testing*. Washington, D.C.: AERA.
- Anastasi, A. (1976). *Psychological testing* (4<sup>th</sup> ed.). New York, NY: Macmillan.
- Anastasi, A. (1988). *Psychological testing* (6<sup>th</sup> ed.). New York, NY: Macmillan.
- Anderman, L. H., & Wolters, C. (2006). Goal, values, and affect: Influences on student motivation. In P. Alexander & P. Winne (Eds.), *Handbook of educational psychology* (2<sup>nd</sup> ed., pp. 369–389). Mahwah, NJ: Erlbaum.
- Anderson, G. L., Herr, K., & Nihlen, A. S. (1994). *Studying your own school: An educator's guide to qualitative practitioner research*. Thousand Oaks, CA: Sage.
- Anderson, G., & Herr, K. (1999). The new paradigm wars: Is there room for rigorous practitioner knowledge in schools and universities? *Educational Researcher*, 28(5), 12–21.
- Andrews, P. (2007). Negotiating meaning in cross-national studies of mathematics teaching: Kissing frogs to find princes. *Comparative Education*, 43, 489–509.

- Ariani, M. G., & GhafouRnia, N. (2015). The relationship between socio-economic status and beliefs about language learning: A study of Iranian postgraduate EAP students. *English Language Teaching*, (8)9, 17–25.
- Arum, R., & Roksa, J. (2011). *Academically adrift: Limited learning on college campuses*. Chicago, IL: University of Chicago Press.
- Bachman, L. (1990). *Fundamental Considerations in Language Testing*. Oxford, England: Oxford University Press.
- Bachman, L. F. (2000). Modern language testing at the turn of the century: Assuring that what we count counts. *Language Testing*, 17, 1–42.
- Bachman, L. F. (2004). *Statistical analyses for language assessment*. Cambridge, United Kingdom: Cambridge University Press.
- Bachman, L. F. (2007). What is the construct? The dialectic of abilities and contexts in defining constructs in language assessment. In J. Fox, M. Wesche, D. Bayliss, L. Cheng, C. E. Turner, & C. Doe (Eds.), *Language testing reconsidered* (pp. 41–71). Ottawa, Canada: University of Ottawa Press.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford, England: Oxford University Press.
- Bachman, L., & Palmer, A. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. New York, NY: Oxford University Press.
- Badawood, O. (2003). *The development of leadership skills of private high school principals in Jeddah, Saudi Arabia* (Unpublished doctoral dissertation). University of Pittsburgh, Pittsburgh, PA.
- Bailey, K. M. (1998). *Learning about language assessment: Dilemmas, decisions, and directions*. Boston, MA: Heinle & Heinle.
- Baker, D. (1989). *Language testing: A critical survey and practical guide*. London, England: Edward Arnold.
- Baker, W. (2011). From cultural awareness to intercultural awareness: culture in ELT. *ELT Journal*, 66(1), 62–70.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: Freeman.
- Barbour, R. (2007). *Doing focus groups*. London, England: SAGE Publications.

- Barnes, N., Fives, H., & Dacey, C. M. (2015). Teachers' beliefs about assessment. In H. Fives & M. G. Gill (Eds.), *International handbook of research on teacher beliefs* (pp. 284–300). New York, NY: Routledge.
- Bearne, E. (1996). *Differentiation and diversity in the primary school*. London, England: Routledge.
- Bell, J. (1999). *Doing your research project* (3<sup>rd</sup> ed.). Buckingham, England: Open University Press.
- Berg, B. L. (2007). *Qualitative research methods for the social sciences*. London, England: Pearson.
- Bernhardt, E. B., Rivera, R. J., & Kamil, M. L. (2004). The practicality and efficiency of web-based placement testing for college-level language programs. *Foreign Language Annals*, 37(3), 356–366.
- Berthold, M. (n.d.). *Reliability of quick placement test: How much faith can we place on quick paper or internet based placement tests?* Retrieved from <http://humanities.curtin.edu.au/schools/EDU/education/pdf/alaa/M-Berthold.pdf>
- Biernat, M., & Dovidio, J. F. (2000). Stigma and stereotypes. In T. F. Heatherton, R. E. Kleck, M. R. Hebl, & J. G. Hull (Eds.), *The social psychology of stigma* (pp. 88–125). New York, NY: Guilford.
- Birnie, B. F. (2015). Making the case for differentiation. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 88(2), 62–65.
- Birch, S. H., & Ladd, G. W. (1996). Interpersonal relationships in the school environment and children's early school adjustment: the role of teachers and peers. In J. Juvonen, & K. Wentzel (Eds.), *Social motivation: Understanding children's school adjustment*. New York: Cambridge University Press.
- Black, P., & Wiliam, D. (2006). Developing a theory of formative assessment. In J. Gardner (Ed.), *Assessment and learning* (pp. 81–100). Thousand Oaks, CA: Sage.
- Black, T. R. (1999). *Doing quantitative research in the social sciences: An integrated approach to research design, measurement and statistics*. Thousand Oaks, CA: Sage.
- Bloomfield, L. (1933/2001). *Language*. Beijing, China: Foreign Language Teaching and Research Press.

- Bloxham, S. (2012). 'You can see the quality in front of your eyes': Grounding academic standards between rationality and interpretation. *Quality in Higher Education*, 18(2), 185–204.
- Bloxham, S., Boyd, P., & Orr, S. (2011). Mark my words: The role of assessment criteria in UK higher education grading practices. *Studies in Higher Education*, 36(6), 655–670.
- Bond, T., & Fox, C. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2<sup>nd</sup> ed.). Mahwah, NJ: LEA.
- Bonner, A., & Tolhurst, G. (2002). Insider/outsider perspectives of participant observation. *Nurse Researcher*, 9(4), 7–19.
- Bonner, S. M. (2016). Teachers' perceptions about assessment: Competing narratives. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of human and social conditions in assessment* (New York, NY: Routledge), 21–39.
- Borg, M. (2001). Key concepts in ELT in Teacher's Beliefs. *ELT Journals*, 55(2), 186–188.
- Borg, S. (1999). Studying teacher cognition in second language grammar teaching. *System* 27, 19–31.
- Borg, S. (2003). Teacher cognition in language teaching: A review of research on what teachers think, know, believe, and do. *Language Teaching*, 36, 81–109.
- Borg, S. (2015). *Teacher cognition and language education: Research and practice*. London, England: Bloomsbury.
- Boyatzis, R. E. (1998). *Transforming qualitative information: Thematic analysis and code development*. London, England: Sage.
- Boyd, A. M. (2003). *Strategies for controlling testlet exposure rates in computerized adaptive testing systems* (Unpublished doctoral thesis). The University of Texas at Austin, Austin, TX.
- Brannen, J. (2005). Mixing methods: The entry of qualitative and quantitative approaches into the research process. *The International Journal of Social Research Methodology, Special Issue*, 8(3), 173–185.
- Brannick, T., & Coghlan, D. (2007). In defense of being "native": The case for insider academic research. *Organizational Research Methods*, 10(1), 59–74.  
<http://dx.doi.org/10.1177/1094428106289253>

- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3, 77–101.
- Breen, M. P., Hird, B., Milton, M., Oliver, R., & Thwaite, A. (2001). Making sense of language teaching: Teachers' principles and classroom practices. *Applied Linguistics* 22(4), 470–501. doi: 10.1093/applin/22.4.470
- Brekhus, W. (1998). A sociology of the unmarked: Redirecting our focus. *Sociological Theory*, 16(1), 34–51.
- British Education Research Association. (2011). *Ethical guidelines for educational research*. London, England: Author. Retrieved from <https://www.bera.ac.uk/wp-content/uploads/2014/02/BERA-Ethical-Guidelines-2011.pdf?noredirect=1>
- Brooke, K., Aden, M., Al-Kuwari, N., Christopher, V., Ibrahim, M., Johnson, B., & Souya, O. (2012). Placement testing in EFL context. *TESOL Arabia Perspective* 19(2), 13–20.
- Brookhart, S. M. (2002). What will teachers know about assessment, and how will that improve instruction. In R. W. Lizzitz & W. D. Schafer (Eds.), *Assessment in educational reform: Both means and ends* (pp. 2–17). Boston, MA: Allyn & Bacon.
- Brown, D. (1983). Conversational cloze tests and conversational ability. *English Language Teaching Journal*, 37(2), 158–161.
- Brown, G. T. L. (2008). *Conceptions of assessment: Understanding what assessment means to teachers and students*. New York, NY: Nova Science Publishers.
- Brown, G. T. L., & Harris, L. R. (Eds.). (2016). *Handbook of social and human conditions in assessment*. New York, NY: Routledge.
- Brown, J. D. (1996). *Testing in language programs*. Upper Saddle River, NJ: Prentice-Hall Regents.
- Brown, J. D. (2002). Extraneous variables and the washback effect. *JALT* 6(2), 12–15.
- Brown, J. D. (2014). *Mixed methods research for TESOL*. Edinburgh, Scotland: University of Edinburgh Press.
- Brumfit, C. J. (2001). *Individual Freedom in Language Teaching: Language Education and Applied Linguistics*. Oxford: Oxford University Press.
- Byram, M. (1989). *Cultural Studies in Foreign Language Education*. Clevedon: Multilingual Matters.

- Byram, M. (1997) *Teaching and Assessing Intercultural Communicative Competence*. Clevedon: Multilingual Matters.
- Byram, M. and Morgan, C. (1994) *Teaching-and-Learning-Language-and-Culture*. Clevedon: Multilingual Matters.
- Byram, M., & Grundy, P. (2003). *Context and culture in language teaching and learning*. Clevedon: Multilingual Matters.
- Byram, M. (2009) Intercultural Competence in Foreign Languages: The Intercultural Speaker and the Pedagogy of Foreign Language Education. In D. K. Deardoff (ed.) *The SAGE Handbook of Intercultural Competence*. Thousand Oaks, CA: SAGE.
- Bryman, A. (2001). *Social research methods*. Oxford, England: Oxford University Press.
- Bryman, A. (2004). *Social research methods* (2<sup>nd</sup> ed.). Oxford, England: Oxford University Press.
- Bryman, A. (2008). *Social research methods* (3<sup>rd</sup> ed.). New York, NY: Oxford University Press.
- Bryman, A. (2012). *Social research methods* (4<sup>th</sup> ed.). Oxford, England: Oxford University Press.
- Bryman, A. (2016). *Social research methods* (5<sup>th</sup> ed.). London, England: Oxford University Press.
- Bryman, A., & Bell, E. (2003). *Business research methods*. Oxford, England: Oxford University Press.
- Bunderson, C. V., Inouye, D. K., & Olsen, J. B. (1989). The four generations of computerized educational measurement. In R. L. Linn (Ed.), *Educational measurement* (pp. 367–407). Washington, DC: American Council on Education.
- Burns, A. (1992). Teacher beliefs and their influence on classroom practice. *Prospect*, 7(3), 56–66.
- Busch, D. (2010). Pre-service teacher beliefs about language learning: The second language acquisition course as an agent for change. *Language Teaching Research*, 14, 318–337.
- Butterworth, D. (2010). *Placing gifted students at risk in mixed-ability classroom: A sequential mixed methods analysis*. Retrieved from ProQuest Dissertations and Theses.

- Byram, M. (2009). Intercultural competence in foreign languages: The intercultural speaker and the pedagogy of foreign language education. In D. K. Deardoff (Ed.), *The SAGE handbook of intercultural competence*. Thousand Oaks, CA: SAGE.
- Byrne, B. (2004). Qualitative interviewing. In C. Seale (Ed.), *Researching society and culture* (2<sup>nd</sup> ed.). London, England: Sage Publications.
- Calderhead, J. (1995). Teachers: Beliefs and knowledge. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology*. New York, NY: Simon & Schuster, Macmillan.
- Callahan, S. (1993, November). *Mathematics placement at Cottey College*. Paper presented at the annual conference of the American Mathematical Association of Two-Year Colleges, Boston, MA. (ERIC Document Reproduction Service No. ED 373813).
- Camilli, G. (2013). Ongoing issues in test fairness. *Educational Research and Evaluation*, 19(2–3), 104–120.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81–105.
- Carroll, John B. (1967). Foreign language proficiency levels attained by language majors near graduation from college. *Foreign Language Annals* 1.131-51.
- Carr, E. C. J., & Worth, A. (2001). The use of the telephone interview for research. *Nursing Times Research*, 6(3), 511–524.
- Carr, D. (2005). Personal and interpersonal relationships in education and teaching: a virtual ethical perspective. *British Journal of Educational Studies*, 53(3), 255–271.
- Casazza, M., & Silverman, S. (1996). *Learning assistance and developmental education*. San Francisco, CA: Jossey-Bass.
- Chang, D. F., & Sue, S. (2003). The effects of race and problem type on teachers' assessments of student behaviour. *Journal of Consulting and Clinical Psychology*, 71, 235–242.
- Chapelle, C. A., & Douglas, D. (2006). *Assessing language through computer technology*. Cambridge, United Kingdom: Cambridge University Press.
- Chapelle, C., & Duff, P. (Eds.). (2003). Some guidelines for conducting quantitative and qualitative research in TESOL. *TESOL Quarterly*, 37, 157–178.
- Chapple, A. (1999). The use of telephone interviewing for qualitative research. *Nurse Researcher*, 6(3), 85–93.

- Chase, C. I. (1999). *Contemporary assessment for educators*. New York, NY: Longman.
- Chenail, R. (2008). *Case studies in 'The Sage encyclopedia of qualitative research methods'*. L. M. Given (Ed.). Thousand Oaks, CA: Sage Publications, Inc.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, Massachusetts: The M.I.T. Press.
- Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education* (6<sup>th</sup> ed.). New York, NY: Routledge.
- Cohen, L., Manion, L., & Morrison, K. (2011). *Research methods in education* (7<sup>th</sup> ed.). New York, NY: Routledge.
- Collie, R. J., Shapka, J. D., & Perry, N. E. (2012). School climate and social–emotional learning: Predicting teacher stress, job satisfaction, and teaching efficacy. *Journal of Educational Psychology*, 104(4), 1189–1204. doi: 10.1037/a0029356
- Coombe, C., Folse, K., & Hubley, N. (2007). *A practical guide to assessing English language learners*. Ann Arbor, MI: University of Michigan Press.
- Cooper, C. R., & Denner, J. (1998). Theories linking culture and psychology: Universal and community-specific processes. *Annual Review of Psychology*, 49, 559–584.
- Corbin, J., & Strauss, A. (2015). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (4<sup>th</sup> ed.). Thousand Oaks, CA: Sage.
- Costley, C., & Gibbs, P. (2006). Researching others: Care as an ethic for practitioner researchers. *Studies in Higher Education*, 31(1), 89–98.  
DOI:10.1080/03075070500392375
- Cothran, D. J., & Ennis, C. D. (1997). Students' and teachers' perceptions of conflict and power. *Teaching and Teacher Education*, 13(5), 541–553.
- Council of Europe. (2001). *Common European Framework of Reference for Languages: Language, teaching, assessment*. Cambridge, United Kingdom: Cambridge University Press.
- Council of Europe. (2009). *Manual for relating language examinations to the Common European Framework of Reference for Languages: Learning, teaching, assessment*. Strasbourg: Council of Europe, Language Policy Division.
- Council of Europe. (2015). *Council of Europe in brief*. Retrieved from <http://www.coe.int/en/web/about-us/who-we-are>



- Creswell, J. W. (2003). *Research design: Qualitative, quantitative, and mixed methods approaches* (2<sup>nd</sup> ed.). Thousand Oaks, CA: Sage.
- Creswell, J. W. (2009). Mapping the field of mixed methods research. *Journal of Mixed Methods Research*, 3(2), 95–108.
- Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4<sup>th</sup> ed.). Thousand Oaks, CA: Sage.
- Creswell, J. W., & Plano Clark, V. L. (2007). *Designing and conducting mixed methods research*. Thousand Oaks, CA: Sage.
- Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research* (2<sup>nd</sup> ed.). Thousand Oaks, CA: Sage Publications, Inc.
- Crnic, K., & Lamberty, G. (1994). Reconsidering school readiness: Conceptual and applied perspectives. *Early Education and Development*, 5(2), 99–105.
- Crocker, L. (2005). Teaching for the test: How and why test preparation is appropriate. In R. Phelps (Ed.), *Defending standardized testing* (pp. 159–174). Mahwah, NJ: Lawrence Erlbaum.
- Crocker, L. (2006). Preparing examinees for test taking: Guidelines for test developers and test users. In S. Downing & T. Haladyna (Eds.), *Handbook of test development* (pp. 115–128). Mahwah, NJ: Lawrence Erlbaum.
- Cronbach, L.J. (1949). *Essentials of psychological testing* (2nd ed.). Oxford, England: Harper.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281-302.
- Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.). *Educational measurement* (2<sup>nd</sup> ed.). Washington, DC: American Council on Education.
- Crozet, C., & Liddicoat, A.J. (1999). The challenge of intercultural language teaching: Engaging with culture in the classroom. J. Lo Bianco, A.J. Liddicoat & C. Crozet (eds), *Striving for the Third Place: Intercultural competence through language education* (pp. 113-126). Canberra: Language Australia.
- Crozet, C. (2016). On language and interculturality: Teaching languages and cultures for a global world' in Julian C.H Lee (ed.) *Narratives of Globalization Reflection on the Global Condition*, Rowman and Littlefield, London, United Kingdom, pp. 85-94.

- Crusan, D. (2014). Placement testing. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 17–25). Hoboken, NJ: Wiley/Blackwell.
- Davis, H. A. (2003). Conceptualizing the role and influence of student–teacher relationships on children’s social and cognitive development. *Educational Psychologist*, 38(4), 207–234.
- Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York, NY: Plenum.
- Deci, E. L., & Ryan, R. M. (2000). The “what” and “why” of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11, 227–268.
- DeLyser, D. (2001). "Do you really live here?" Thoughts on insider research. *Geographical Review*, 91(1/2), 441–453.
- Denscombe, M. (2014). *The good research guide: For small-scale social research projects*. (5<sup>th</sup> ed.). Maidenhead, England: Open University Press.
- Denzin, N. (2008). The new paradigm dialog and qualitative inquiry. *International Journal of Qualitative Studies in Education*, 21(4), 315–325.
- Denzin, N. K. (1970). *The research act*. Chicago, IL: Aldine.
- Denzin, N. K. (1989). *Interpretive interactionism*. Newbury Park, CA: Sage.
- Denzin, N. K., & Lincoln, Y. S. (2005). Introduction: The discipline and practice of qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research* (2<sup>nd</sup> ed.). Thousand Oaks, CA: Sage.
- Deraney, P. (2015). Analysis of systemic functional linguistics preparatory year writing in a Saudi university. In A. K. Hamdan (Ed.), *Teaching and learning in Saudi Arabia*. Rotterdam, Netherlands: Sense Publishers.
- Dörnyei, Z. (1994). Understanding L2 motivation: On with the challenge! *Modern Language Journal*, 78, 515–523.
- Dörnyei, Z. (2001). *Motivational strategies in the language classroom*. Cambridge, United Kingdom: Cambridge University Press.
- Dörnyei, Z. (2007). *Research methods in applied linguistics*. New York, NY: Oxford University Press.
- Dörnyei, Z., & Csizér, K. (1998). Ten commandments for motivating language learners: Results of an empirical study. *Language Teaching Research*, 2, 203–229.

- Dörnyei, Z., & Ushioda, E. (2013). Theories of motivation in psychology. In Z. Dörnyei & E. Ushioda (Eds.), *Teaching and researching motivation* (pp. 12–38). Old Tappan, NJ: Taylor and Francis.
- Douglas, D., & Hegelheimer, V. (2007). Assessing language using computer technology. *Annual Review of Applied Linguistics*, 27, 115–132.
- Dovidio, J. F., Major, B., & Crocker, J. (2000). Stigma: Introduction and overview. In T. F. Heatherton, R. E. Kleck, M. R. Hebl, & J. G. Hull (Eds.), *Stigma: Social psychological perspectives* (pp. 1–28). New York, NY: Guilford.
- Dulay, H., & Burt, M. (1977). Remarks on creativity in language acquisition. In M. Burt, H. Dulay, & M. Finnochiaro (Eds.), *Viewpoints on English as a second language* (pp. 95–126). New York, NY: Regents Press.
- Easterby-Smith, M. (2002). *Management research*. London, England: Sage.
- Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, 53, 109–132.
- Eggen, T. J. H. M. (2001). Overexposure and underexposure of items in computerized adaptive testing. *Measurement and Research Department Reports 2001-1*. Arnhem, Netherlands: Citogroep.
- Elliot, N. (2005). *On a scale: A social history of writing assessment in America*. York, NY: Peter Lang.
- Ellis, R. (2006). Current issues in the teaching of grammar: An SLA perspective. *TESOL Quarterly*, 40(1), 83–107.
- Elyas, T., & Al Grigri, W. H. (2014). Obstacles to teaching English in Saudi Arabia public schools: Teachers' and supervisors' perceptions. *International Journal of Language Teaching*, 2(3), 74–89.
- Engelsen, K. S., & Smith, K. (2014). Assessment literacy. In C. Wyatt-Smith, V. Klenowski, & P. Colbert (Eds.), *The enabling power of assessment: Designing assessment for quality learning* (pp. 140–162). New York, NY: Springer.
- English Language Institute. (2017). *ELI Student Handbook 2017/2018*. Retrieved from [www.eli.kau.edu.sa](http://www.eli.kau.edu.sa)
- Ercikan, K., & Roth, W. (2006). What good is polarizing research into qualitative and quantitative? *Educational Researcher*, 35, 14–23.

- Ercikan, K., & Solano-Flores, G. (2016). Assessment and sociocultural context: A bidirectional relationship. In G. T. L. Brown & L. Harris (Eds.), *Human factors and social conditions of assessment*. New York, NY: Routledge.
- Eren, A. (2010). Consonance and dissonance between Turkish prospective teachers' values and practices: Conceptions about teaching, learning, and assessment. *Australian Journal of Teacher Education*, 35(3), 27–48.
- Erickson, F. (2004). Culture in society and in educational practices. In J. A. Banks & C. A. M. Banks (Eds.), *Multicultural education: Issues and perspectives* (5th ed., pp. 31–55). Hoboken, NJ: John Wiley and Sons.
- Eriksson, P., & Kovalainen, A. (2008). *Qualitative methods in business research*. London, England: Sage.
- Exam English LTD. (2014). Retrieved March 29, 2018 from <https://www.examenglish.com/IELTS/index.html>
- Fan, W. (2011). Social influences, school motivation and gender differences: An application of the expectancy-value theory. *Educational Psychology Review*, 31, 157–175.
- Faruk, S. (2014). Saudis' attitude towards English: Trend and rationale. *Professional Communication and Translation Studies*, 7, 173–180.
- Figueras, N. (2007). The CEFR, a lever for the improvement of language professionals in Europe. *Modern Language Journal*, 91, 673–675.
- Flick, U. (2014). *An introduction to qualitative research* (5<sup>th</sup> ed.). London, England: SAGE.
- Fox, J. (2009). Moderating top-down policy impact and supporting EAP curricular renewal: Exploring the potential of diagnostic assessment. *Journal of English for Academic Purposes*, 8, 26–42.
- Freed, B. F., Dewey, D. P., Segalowitz, N., & Halter, R. (2004). The language contact profile. *Studies in Second Language Acquisition*, 26(2), 349–356.
- Fulcher, G. (1999). Ethics in language testing. *TAE SIG Newsletter*, 1(1), 1–4.
- Fulcher, G. (2004). Deluded by artifices? The Common European Framework and harmonization. *Language Assessment Quarterly*, 1, 253–66.

- Fulcher, G. (2010). *Practical language testing*. London, England: Hodder Education.
- Fulcher, G., & Davidson, F. (2007). *Language testing and assessment: An advanced resource book*. London, England: Routledge.
- Fulmer, G. W., Lee, I. C. H., & Tan, K. H. K. (2015). Multi-level model of contextual factors and teachers' assessment practices: An integrative review of research. *Assessment in Education: Principles, Policy, and Practice*, 22, 475–494. doi:10.1080/0969594X.2015.1017445
- Gage, N. L. (1989). The paradigm wars and their aftermath. *Educational Researcher*, 18(7), 4–10.
- Gardner, R. C. (1985). *Social psychology and second language learning: The role of attitudes and motivation*. London, England: Edward Arnold Publishers.
- Gebril, A. (2006). *Independent and Integrated academic writing tasks: A study in generalizability and test method* (Unpublished doctoral dissertation). The University of Iowa, Iowa City.
- Gerrish, K. (1997). Being a 'marginal native': Dilemmas of the participant observer. *Nurse Researcher*, 5(1), 25–34.
- Gibbs, P., & Costley, C. (2006). An ethics of community and care for practitioner researchers. *International Journal of Research and Method in Education*, 29(2), 239–249.
- Gierl, M. J., Lai, H., & Li, J. (2013). Identifying differential item functioning in multi-stage computer adaptive testing. *Educational Research and Evaluation*, 18(2–3), 188–203.
- Gipps, C. (2012). *Beyond testing: Towards a theory of educational assessment*. Abingdon, England: Routledge.
- Gipps, C., & Murphy, P. (1994). *A fair test? Assessment, achievement and equity*. Buckingham, England: Open University Press.
- Given, M. (2008). *The Sage encyclopedia of qualitative research methods*. Thousand Oaks, CA: Sage.
- Glas, K., & Cárdenas-Claros, M. (2013). New English cultures and learner autonomy for intrinsic motivation and democratic empowerment in the Chilean context. *English Teaching: Practice Critique*, 12, 23–40.

- Goffman, E. (1963). *Stigma: Notes on the management of spoiled identity*. New York, NY: Prentice Hall.
- Gorard, S., & Smith, E. (2006). Beyond the 'learning society': What have we learnt from widening participation research? *International Journal of Lifelong Education*, 25(6), 575–594.
- Gorard, S., & Taylor, C. (2004). What is 'triangulation'? *Building research capacity* (Vol. 7, pp. 7–9). Cardiff, Wales: Cardiff University, Cardiff School of Social Sciences.
- Goullier, F. (2006). *Council of Europe tools for language teaching: Common European Framework and portfolios*. Retrieved from [http://www.coe.int/t/dg4/linguistic/Source/Goullier\\_Outils\\_EN.pdf](http://www.coe.int/t/dg4/linguistic/Source/Goullier_Outils_EN.pdf)
- Gray, J. (2000) 'The ELT Coursebook as Cultural Artefact: how teachers censor and adapt'. ELT
- Green, A. (2012). Placement testing. In C. Coombe, P. Davidson, B. O'Sullivan, & S. Stoyneff (Eds.), *The Cambridge guide to second language assessment* (pp.164–170). New York, NY: Cambridge University Press.
- Green, A. (2014). *Exploring language assessment and testing: Language in action*. Oxon, United Kingdom: Routledge.
- Green, A. B., & Weir, C. J. (2004). Can placement tests inform instructional decisions? *Language Testing*, 22(4), 467–494.
- Greene, B. A., Miller, R. B., Crowson, H. M., Duke, B. L., & Akey, K. L. (2004). Predicting high school students' cognitive engagement and achievement: Contributions of classroom perceptions and motivation. *Contemporary Educational Psychology*, 29(4), 462–482.
- Greene, H.A., Jorgensen, A.N. and Gerberich, J.R. (1943). *Measurement and Evaluation in the Secondary School*. New York: Longmans, Green and Co.
- Greenwood, D., & Levin, M. (2007). *Introduction to action research* (2<sup>nd</sup> ed.). Thousand Oaks, CA: Sage Publications.
- Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Newbury Park, CA: Sage.
- Guhn, M., Zumbo, B. D., Janus, M., & Hertzman, C. (2011). Validation theory and research for a population-level measure of children's development, wellbeing, and school readiness. *Social Indicators Research*, 103, 183–191.

- Guilford, J. P. (1946). New standards for test evaluation. *Educational and Psychological Measurement*, 6(4), 427–439.
- Hagler, A. (2014). A study of attitudes toward western culture among Saudi university students. *Learning and Teaching in Higher Education: Gulf Perspectives*, 11(1), 1–14.
- Hallam, S., & Ireson, J. (2005). Secondary school teachers' pedagogic practices when teaching mixed and structured ability classes. *Research Papers in Education*, 20(1), 3–24.
- Hamdan, A. (2015). *Teaching and learning in Saudi Arabia: Perspectives from higher education*. Rotterdam, Netherlands: Sense Publishers.
- Hammersley, M. (1992). *What's wrong with ethnography?: Methodological explorations*. London, England: Routledge.
- Hammersley, M. (Ed.). (1993). *Educational research: Current issues*. London, England: Paul Chapman Publishing.
- Hamp-Lyons, L. (1997). Washback, impact and validity: Ethical concerns. *Language Testing*, 14(3), 295–303.
- Hardre, P. L., & Sullivan, D. W. (2008). Teacher perceptions and individual differences: How they influence teachers' motivating strategies. *Journal of Teaching and Teacher Education*, 24, 2059–2075.
- Hardre, P. L., & Sullivan, D. W. (2009). Motivating adolescents: Teachers' beliefs, perceptions and classroom practices. *Teacher Development*, 13, 1–16.
- Hardre, P. L., Crowson, H. M., Debacker, T. K., & White, D. (2007). Predicting the academic motivation of rural high school students. *Journal of Experimental Education*, 75(4), 247–269.
- Hardre, P. L., Huang, S. H., Chen, C. H., Chiang, C. T., Jen, F. L., & Warden, L. (2006). High school teachers' motivational perceptions and strategies in an East Asian nation. *Asia-Pacific Journal of Teacher Education*, 34, 199–221.
- Harlen, W. (2007). *Assessment of Learning*. London, England: Sage.
- Harrison, A. (1983). *A language testing handbook*. London, England: Macmillan.
- Hassan, N. (2013). The impact of teachers' beliefs on L2 grammar teaching. *Language in India*, 13(8).

- He, Q., & Tymms, P. B. (2005). A computer-assisted test design and diagnosis system for use by classroom teachers. *Journal of Computer Assisted Learning*, 21, 419–429.
- Heaton, J. B. (1990). *Classroom testing*. Harlow, England: Longman.
- Hewitt-Taylor, J. (2002). Insider knowledge: Issues in insider research. *Nursing Standard*, 16(46), 33–35.
- Heyman, G. D., & Compton, B. J. (2006). Context sensitivity in children's reasoning about ability across the elementary school years. *Developmental Science*, 9(6), 616–627.
- Hofstede, G. (2011). Dimensionalizing Cultures: The Hofstede Model in Context. *Online Readings in Psychology and Culture*, 1-26.
- Holliday, A. (2005). *The Struggle to Teach English as an International Language*. Oxford: Oxford University Press.
- Holliday, A. (2016). *Doing & writing qualitative research* (3<sup>rd</sup> ed.). London, England: Sage.
- Holloway, I., & Todres, L. (2003). The status of method: Flexibility, consistency and coherence. *Qualitative Research*, 3(3), 345–357.
- Holmes, P., Fay, R., Andrews, J., & Attia, M. (2013). Researching multilingually: New theoretical and methodological directions. *International Journal of Applied Linguistics*, 23(3), 285–299.
- Hoshmand, L. T. (2003). Can lessons of history and logical analysis ensure progress in psychological science? *Theory and Psychology*, 13, 39–44.
- Huang, Y.-M., Lin, Y.-T., & Cheng, S.-C. (2009). An adaptive testing system for supporting versatile educational assessment. *Computers & Education*, 52(1), 53–67.
- Hubley, A. M., & Zumbo, B. D. (2011). Validity and the consequences of test interpretation and use. *Social Indicators Research*, 103(2), 219-230.
- Hughes, A. (1989). *Testing for language teachers*. Cambridge, United Kingdom: Cambridge University Press.
- Hughes, A. (1994). *Backwash and TOEFL 2000*. Unpublished manuscript, University of Reading, England.



- Hughes, A. (2003). *Testing for Language Teachers* (2<sup>nd</sup> ed.). Cambridge, United Kingdom: Cambridge University Press.
- Hughes, K. L., & Scott-Clayton, J. (2011). *Assessing developmental assessment in community colleges* (CCRC Working Paper No. 19, Assessment of Evidence Series). New York, NY: Columbia University, Teachers College, Community College Research Center.
- Hulse, S. H., Egeth, H., & Deese, J. (1981). *The Psychology of Learning* (5<sup>th</sup> ed.). Johannesburg, South Africa: McGraw-Hill International Book Company.
- Hulstijn, J. (2007). The shaky ground beneath the CEFR: Quantitative and qualitative dimensions of language proficiency. *Modern Language Journal*, 91(4), 663–667.
- Hussain, M. P., Albeshier, K. B., & Farid, A. (2017). Preparing preparatory year English language learners for professional colleges: An evaluation of current placement system. *Arab World English Journal*, 8(2), 36–51.
- Ichado, S. M. (1998). Impact of broken home on academic performance of secondary school students in English Language. *Journal of Research in Counseling Psychology*, 4(1), 84–87.
- Iliescu, D. (2017). Equivalence in Test Adaptations. In *Adapting Tests in Linguistic and Cultural Situations* (Educational and Psychological Testing in a Global Context, pp. 79-124). Cambridge: Cambridge University Press.
- Ingram, E. (1977). Basic concepts in testing. In J. P. B. Allen & A. Davies (Eds.), *Testing and experimental methods*. Oxford, England: Oxford University Press.
- Ireson, J., & Hallam, S. (2001). *Ability grouping in education*. London, England: Sage.
- Isonio, S. (1992). *Mid-term assessment of English 10 students: A comparison of methods of entry into the course*. Huntington Beach, CA: Golden West College.
- Jamalzadeh, M., & Shahsavar, Z. (2015). The effects of contextual factors on teacher's beliefs and practices. *Procedia – Social and Behavioral Sciences*, 192, 166–171.
- James, N., & Busher, H. (2006). Credibility, authenticity and voice: Dilemmas in web-based interviewing. *Qualitative Research*, 6(4), 403–420.
- Javid, C. Z. (2014). Measuring language anxiety in an EFL Context. *Journal of Education and Practice*, 5(25), 180–192.
- Jenkins, J. (2000). *The phonology of English as an international language: new models, new norms, new goals*. Oxford: Oxford University Press.

- Johnson, K. E. (1992). The relationship between teachers' beliefs and practices during literacy instruction for non-native speakers of English. *Journal of Literacy Research*, 24(1), 83–108. doi: 10.1080/10862969209547763
- Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research*, 1, 112–133. <http://dx.doi.org/10.1177/1558689806298224>
- Jones, S. (1985). The analysis of depth interviews. In R. Walker (Ed.), *Applied qualitative research*. Aldershot, England: Gower.
- Jose, C. P. (2002). *Ingles Para Quimica E Ingenieria Quimica*. Barcelona, Spain: Ariel.
- Jue, P. Y. (1993, March). *Course enrolments and subsequent success after being advised of or 'Blind' to assessment test scores and recommendations*. Napa Valley, CA: Napa Valley College.
- Kagan, D. M. (1992). Implications for research on teacher beliefs. *Educational Psychologist*, 27, 65–90. doi:10.1207/s15326985ep2701\_6
- Kalaja, P., Barcelos, A. M. F., Aro, M., & Ruohotie-Lyhty, M. (2016). *Beliefs, agency and identity in foreign language learning and teaching*. Basingstoke, England: Palgrave.
- Kane, M. (2006). Validation. In Brennan, R. (Ed.), *Educational measurement* (4<sup>th</sup> ed., pp. 17–64). Westport, CT: American Council on Education and Praeger.
- Kaplan, A., & Maehr, M. (2007). The contributions and prospects of goal orientation theory. *Educational Psychology Review*, 19, 141–184.
- Kappler, K. (2004). NCATE: Wolf in Shepard's clothes. In T. Poetter, T. Goodney, & J. Bird (Eds.), *Critical perspectives on the curriculum of teacher education*. Lanham, MD: University Press of America.
- Katz, I. R., & Gorin, J. S. (2016). Computerising assessment: Impacts on education stakeholders. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of human and social conditions in assessment* (pp. 472–489). New York, NY: Routledge.
- Kennedy, J. (1996). The role of teacher beliefs and attitudes in teacher behaviour. In G. T. Sachs, M. Brock, & R. Lo (Eds.), *Directions in second language teacher education* (pp. 107–122). Hong Kong, China: City University of Hong Kong.
- Kenny, D. (2004). Social desirability bias. In M. S. Lewis-Beck, A. Bryman, & T. F. Liao (Eds.), *The SAGE encyclopedia of social science research methods*. Thousand Oaks, CA: SAGE Publications.

- Kerlinger, F. N. (1973). *Foundations of behavioral research*. New York, NY: Holt, Rinehart and Winston.
- Kerry, T., & Kerry, C. (1997). Differentiation: Teachers' views of the usefulness of recommended strategies in helping the more able pupils in primary and secondary schools. *Educational Studies*, 23(3), 439–457.
- Khan, I. (2011). The teacher of English: Pedagogic relevance in Saudi Arabia. *English Language Teaching*, 4(2), 112–120.
- Khan, M. (2015). Analyzing the Relationship between L2 Motivational Selves and L2 Achievement: A Saudi Perspective. *International Journal of English Language Teaching*, 2(1), 68-75. doi: 10.5430/ijelt.v2n1p68
- Khondker, H. (2004). Glocalization as globalization: Evolution of a sociological concept. *Bangladesh e-Journal of Sociology*, 1(2), 1-9.
- Kim, K. H., & Zabelina, D. (2015). Cultural Bias in Assessment: Can Creativity Assessment Help? *The International Journal of Critical Pedagogy*, 6(2), 129-148.
- King, N. (2004). Template analysis. In C. M. Cassell & G. Symon (Eds.), *Essential guide to qualitative methods in organizational research*. London, England: Sage Publications.
- Kirk, J., & Miller, M. L. (1986). *Reliability and validity in qualitative research*. Beverly Hills, CA: Sage Publications.
- Kokhan, K. (2012). Investigating the possibility of using TOEFL scores for university ESL decision-making: Placement trends and effect of time lag. *Language Testing*, 29(2), 291–308.
- Kokhan, K. (2013). An argument against using standardized test scores for placement of international undergraduate students in English as a Second Language (ESL) courses. *Language Testing*, 30(4), 467–489.
- Kokhan, K., & Lin, C. K. (2014). Test of English as a Foreign Language (TOEFL): Interpretation of multiple score reports for ESL placement. *Papers in Language Testing and Assessment*, 3, 1–23.
- Koutselini, M. (2006). *Diáforopoiisi Didaskalias se taxis Miktis Ikanotitas* [Differentiation of teaching and learning in mixed-ability classrooms]. Nicosia.
- Koutselini, M. (2008). Listening to students' voices for teaching in mixed ability classrooms: Presuppositions and considerations for differentiated instruction. *Learning and Teaching*, 1(1), 17–30.

- Krashen, S. (1985). *The input hypothesis: Issues and implications*. Harlow, England: Longman.
- Kramsch, C. (1993) *Context and Culture in Language Teaching*. Oxford: Oxford University Press.
- Kramsch, C. (2009). *The multilingual subject: What language learners say about their experience and why it matters*. Oxford: Oxford Press.
- Kramsch, C. & Yin, P. (2018). Teaching Foreign Languages in the Glocal Contact Zone: The Case of France and China. In A. F. Selvi & N. Rudolph (Eds.), *Conceptual shifts and contextualised practices in education for global interaction, issues and implications* (pp.17-36). Singapore: Springer.
- Kubanyiova, M. (2006). Developing a motivational teaching practice in EFL teachers in Slovakia: Challenges of promoting teacher change in EFL contexts [Electronic version]. *TESL-EJ special issue: Language education research in international contexts*, 10, A-5.
- Kuhn, T. S. (1970). *The structure of scientific revolutions*. Chicago, IL: University of Chicago Press.
- Kunnan, A. J. (2000). Fairness and justice for all. In A. J. Kunnan (Ed.), *Fairness and validation in language assessment* (pp. 1–14). Cambridge, United Kingdom: Cambridge University Press.
- Kuppens, A. H. (2010). Incidental foreign language acquisition from media exposure. *Learning Media and Technology*, 35(1), 65–85.
- Laar, B. (1995). 'An inspector writes': Times Educational Supplement. 8.12.95
- Lado, R. (1961). *Language testing: The construction and use of foreign language tests: A teacher's book*. Bristol, England: Longmans, Green and Company.
- Lai, E. R., & Waltman, K. (2008). Test preparation: Examining teacher perceptions and practices. *Educational Measurement: Issues and Practice*, 27(2), 28–45.
- Lambert, D., & Lines, D. (2000). *Understanding assessment*. London, England: Routledge Falmer.
- Language Policy Division. (2006). *Plurilingual education in Europe: 50 years of international cooperation*. Retrieved from [http://www.coe.int/t/dg4/linguistic/Source/PlurinlingalEducation\\_En.pdf](http://www.coe.int/t/dg4/linguistic/Source/PlurinlingalEducation_En.pdf)

- Latterell, C. M., & Regal, R. R. (2003). Are placement tests for incoming undergraduate mathematics students worth the expense of administration? *PRIMUS*, 13(2), 152–164. DOI: 10.1080/10511970308984054
- Laurence, S., & Margolis, E. (2012). The scope of the conceptual. In E. Margolis, R. Samuels, & S. P. Stich (Eds.), *The Oxford Handbook of Philosophy of Cognitive Science* (pp. 291–317). Oxford, England: Oxford University Press.
- Lawrence, I. (2011). Reliability and comparability of TOEFL iBT scores. *TOEFL iBT Research Series*, 1(3).
- Leeson, H. (2006). The mode effect: A literature review of human and technological issues in computerized testing. *International Journal of Testing*, 6(1), 1–24.
- Lepper, M. R., & Greene, D. (1978). *The hidden costs of reward*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Levin, T., & Wadmany, R. (2006). Listening to students' voices on learning with information technologies in a rich technology-based classroom. *Journal of Educational Computing Research*, 34(3), 295–331.
- Li, L. (2013). The complexity of language teachers' beliefs and practice: One EFL teacher's theories. *The Language Learning Journal*, 41(2), 175–191.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Newbury Park, CA: Sage Publications.
- Lincoln, Y. S., & Guba, E. G. (2000). Paradigmatic controversies, contradictions, and emerging confluences. In N. K. Denzin & Y. S. Lincoln (Eds.), *The handbook of qualitative research* (2<sup>nd</sup> ed., pp. 163–188). Beverly Hills, CA: Sage.
- Linnenbrink, E. A., & Pintrich, P. R. (2002). Achievement goal theory and affect: An asymmetrical bidirectional model. *Educational Psychologist*, 37(2), 69–78.
- Little, D. (2007). The Common European Framework of Reference for Languages: Perspectives on the making of supranational language education policy. *The Modern Language Journal*, 91(4), 645–655.
- Loiacono, V., & Allen, B. (2008). Are special education teachers prepared to teach the increasing number of students diagnosed with autism? *International Journal of Special Education*, 23, 120–127.
- Long, A. Y., Shin, S.-Y., Geeslin, K., & Willis, E. W. (2018). Does the test work? Evaluating a web-based language placement test. *Language Learning & Technology*, 22(1), 137–156.

- Long, M. H. (1981). Input, interaction, second-language acquisition. In H. Winitz (Ed.), *Native language and foreign language acquisition* (pp.259–278). New York, NY: New York Academy of Sciences.
- Luecht, R. M., Champlain, A. D., & Nungester, R. J. (1998). Maintaining content validity in computerized adaptive testing. *Advances in Health Sciences Education*, 3, 29–41.
- Maddox, B. (2014). Globalising assessment: An ethnography of literacy assessment, camels and fast food in the Mongolian gobi. *Comparative Education*, 50(4), 474–489.
- Maddox, B. (2015). The neglected situation: Assessment performance and interaction in context. *Assessment in Education: Principles, Policy and Practice*, 22(4), 427–443
- Mahmoud, S. (2014). A framework for investigating foundation year student's needs in general English in KAU: Needs analysis extended to curriculum development. *Journal of Language Teaching and Research*, 5(2), 335–342. <http://dx.doi.org/10.4304/jltr.5.2.335-342>
- Marsella, A. J., & Leong, F. T. L. (1995). Cross-cultural issues in personality and career assessment. *Journal of Career Assessment*, 3, 202–218.
- Marshall, J. P., & Allen, B. D. (2000). *The development and implementation of a mathematics placement program*. Retrieved from ERIC database. (ED444850)
- Mattheoudakis, M. (2007). Tracking changes in pre-service EFL teacher beliefs in Greece: A longitudinal study. *Teaching and Teacher Education*, 23, 1272–1288. <http://dx.doi.org/10.1016/j.tate.2006.06.001>
- Maxwell, M. (1997). *Improving student learning skills*. Clearwater, FL: H and H Publishing.
- Maykut, P. S., & Morehouse, R. E. (1994). *Beginning qualitative research: A philosophic and practical guide*. London, England: Falmer Press.
- McCabe, R. (2000). *No one to waste: A report to public decision-makers and community college leaders*. Washington, DC: Community College Press.
- McCoyd, J. L., & Kerson, T. S. (2006). Conducting intensive interviews using email: A serendipitous comparative opportunity. *Qualitative Social Work*, 5, 389–406.
- McMillan, J. H. (2008). *Assessment essentials for standards-based education* (2nd ed.). California: Corwin Press.
- Meccawy, Z. (2010). *Language, culture and identity: A qualitative study of Saudi female*

*students' perceptions of EFL current practices and textbook use* (Unpublished doctoral thesis). Nottingham University, Nottingham, England.

- Medhanie, A. G., Dupuis, D. N., LeBeau, B., Harwell, M. R., & Post, T. R. (2012). The role of the ACCUPLACER mathematics placement test on a student's first college mathematics course. *Educational and Psychological Measurement, 72*(2), 332–351.
- Mercer, J. (2007). The challenges of insider research in educational institutions: Wielding a double-edged sword and resolving delicate dilemmas. *Oxford Review of Education, 33*(1), 1–17. Retrieved from <http://www.jstor.org.ezphost.dur.ac.uk/stable/4618693>
- Mercer, N. (1995). *The guided construction of knowledge: Talk amongst teachers and learners*. Clevedon, England: Multilingual Matters.
- Merrell, C., & Tymms, P. (2007). Identifying reading problems with computer-adaptive assessments. *Journal of Computer Assisted Learning, 23*(1), 27–35.
- Merriam, S. B. (1988). *Case study research in education: A qualitative approach*. San Francisco, CA: Jossey-Bass.
- Merton, R. (1972). Insiders and outsiders: A chapter in the sociology of knowledge. *American Journal of Sociology, 78*(July), 9–47.
- Messick, S. (1998). Test validity: A matter of consequence. *Social Indicators Research, 45*, 35–44.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-104). New York, NY: American Council on education and Macmillan.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' and performances as scientific inquiry into score meaning. *American Psychologist, 50*, 741–749.
- Messick, S. (1996). Validity and washback in language testing. *Language Testing, 13*(3), 241–256.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis* (2<sup>nd</sup> ed.). Thousand Oaks, CA: Sage Publications.
- Miles, M. B., Huberman, A. M., & Saldana, J. (2014). *Qualitative data analysis: A methods sourcebook* (3<sup>rd</sup> ed.). Thousand Oaks, CA: Sage.
- Miller, M. D., Linn, R. L., & Gronlund, N. E. (2009). *Measurement and assessment in teaching* (10<sup>th</sup> ed.). Upper Saddle River, NJ: Pearson Education, Inc.

- Ministry of Education. (2004). *The development of education*. Paper presented at 47th Session of the International Conference on Education, Geneva, Switzerland.
- Mitchell, R. E., & Cawthon, S. W. (2014). Academic achievement: Test bias and fairness. In G. Gertz & P. Boudreault (Eds.), *Deaf studies encyclopedia*. Thousand Oaks, CA: Sage.
- Mohammed, A. (2016). EFL effective factors: Anxiety and motivation and their effect on Saudi college student's achievement. *Arab World English Journal (AWEJ)*, 6(2). <http://dx.doi.org/10.2139/ssrn.2834391>
- Moisander, J., & Valtonen, A. (2006). *Qualitative marketing research: A cultural approach*. London, England: Sage.
- Moloney, M., Dietrich, A., Strickland, O., & Myerburg, S. (2003). Using Internet discussion boards as virtual focus groups. *Advances in Nursing Science*, 26(4), 274–286.
- Morante, E. (1987). A primer on placement testing. In D. Bray & M. Belcher (Eds.), *New directions for community colleges: Issues in student assessment* (pp.55–63). San Francisco, CA: Jossey Bass.
- Morante, E. (1989). Selecting tests and placing students. *Journal of Developmental Education*, 13(2), 2–4, 6.
- Morante, E. A. (2012). Editorial: What do placement tests measure? *Journal of Developmental Education*, 35(3), 28.
- Mori, Y. (1999). Epistemological beliefs and language learning beliefs: What do language learners believe about their learning? *Language Learning*, 49, 377–415.
- Morrison, K. R. B. (1993). *Planning and accomplishing school centred evaluation*. Norfolk, England: Peter Francis Publishers.
- Moskovsky, C., & Alrabai, F. (2009). Intrinsic motivation in Saudi learners of English as a foreign language. *The Open Applied Linguistics Journal*, 2(1), 1–10.
- Muis, K. R., & Foy, M. J. (2010). The effects of teachers' beliefs on elementary students' beliefs, motivation, and achievement in mathematics. In L. D. Bendixen & F. C. Fuecht (Eds.), *Personal epistemology in the classroom: Theory, research, and implications for practice* (pp. 435–469). New York, NY: Cambridge.
- Mullen, A. (2008). *Impact of using a proficiency test as a placement tool: The case of TOEIC* (Doctoral dissertation). Laval University, Quebec, Canada.



- Muller, C., Katz, S. R., & Dance, L. J. (1999). Investing in teaching and learning: Dynamics of the teacher-student relationship from each actor's perspective. *Urban Education*, 34(3), 292–337.
- Munn, P., & Drever, E. (1990). *Using questionnaires in small-scale research: A teachers' guide*. Edinburgh: SCRE Publication.
- Newby, P. (2014). *Research methods for education*. Oxon, England: Routledge.
- Newton, P.E. and Shaw, S.D. (2014). *Validity in Educational and Psychological Assessment*. London: SAGE.
- Nieto, S. (1996). *Affirming diversity: The sociopolitical context of multicultural education* (2<sup>nd</sup> ed.). New York, NY: Longman.
- Nortier, J. (2008). The Moroccan community in the Netherlands. In A. Crees, P. Martin, & N. H. Hornberger (Eds.), *Encyclopedia of language and education: Ecology of language* (Vol. 9, 2<sup>nd</sup> ed., pp. 195–207). New York, NY: Springer.
- O'Connor, H., Madge, C., Shaw, R., & Wellens, J. (2008). Internet-based interviewing. In N. Fielding, R. M. Lee, & G. Blank (Eds.), *The Sage handbook of online research methods* (pp. 271–289). Thousand Oaks, CA: Sage Publications.
- O'Donovan, B., Price, M., & Rust, C. (2008). Developing student understanding of assessment standards: A nested hierarchy of approaches. *Teaching in Higher Education*, 13(2), 205–16.
- O'Sullivan, B., & Green, A. (2011). Test taker characteristics. In L. Taylor (Ed.), *Examining speaking: Research and practice in assessing second language speaking* (pp. 38–39). Cambridge, United Kingdom: CUP and Cambridge ESOL.
- Onwuegbuzie, A. J., & Leech, N. L. (2005). On becoming a pragmatic researcher: The importance of combining quantitative and qualitative research methodologies. *International Journal of Social Research Methodology*, 8, 375–387.
- Oppenheim, A. N. (1992). *Questionnaire design, interviewing and attitude measurement*. London, England: Pinter Publishers.
- Opre, D. (2015). Teachers' conceptions of assessment. *Procedia – Social and Behavioral Sciences*, 209, 229–233.
- Orr, S. (2008). Real or imagined? The shift from norm referencing to criterion referencing. In A. Havnes & L. McDowell (Eds.), *Balancing dilemmas in assessment and learning in contemporary education*. London, England: Routledge.

- Osborn, M. (2003). *A world of difference?: Comparing learners across Europe*. Maidenhead, United Kingdom: Open University Press.
- Oyaid, A. (2009). *Education policy in Saudi Arabia and its relation to secondary school teachers' ICT use, perceptions, and views of the future of ICT in education* (Unpublished doctoral thesis). University of Exeter, Exeter, United Kingdom. Retrieved from <https://ore.exeter.ac.uk/repository/bitstream/handle/10036/69537/OyaidA.doc.pdf?sequence=2>
- Oyserman, D., & Fryberg, S. (2006). The Possible Selves of Diverse Adolescents: Content and Function Across Gender, Race and National Origin. In C. Dunkel & J. Kerpelman (Eds.), *Possible selves: Theory, research and applications* (pp. 17-39). Hauppauge, NY, US: Nova Science Publishers.
- Pallant, J. (2007). *SPSS survival manual—A step by step guide to data analysis using SPSS for windows* (3<sup>rd</sup> ed.). Maidenhead, England: Open University Press.
- Paltridge, B., & Phakiti, A. (2015). *Research methods in applied linguistics: A practical resource*. London, England: Bloomsbury.
- Papageorgiou, S. (2007). *Relating the Trinity College London GESE and ISE exams to the Common European Framework of Reference: Piloting of the Council of Europe draft Manual* (Final project report). Lancaster: Lancaster University.
- Parsons, R. D., Hinson, S. L., & Sardo-Brown, D. (2001). *Educational psychology: A practitioner-researcher model of teaching*. Toronto, Ontario: Wadsworth.
- Pei, Z., & Xing, Y. (2016). Perceptions of world Englishes accents in English phonetics instruction in China. *International Journal of Applied Linguistics & English Literature*, 5(4).
- Pennycook, A. (1994). *The Cultural Politics of English as an International Language*. Essex, Pearson Longman.
- Pepin, B. (2005). Can we compare like with like in comparative educational research? Methodological considerations in crosscultural studies in mathematics education. In B. Hudson & J. Fagner (Eds.), *Researching the teaching and learning of mathematics II* (pp. 39–54). Linz, Austria: Pädagogische Akademie des Bundes.
- Phipps, S., & Borg, S. (2009). Exploring tensions between teachers' grammar teaching beliefs and practices. *System*, 37(3), 380–390.

- Pollit, A. (2009). *The Oxford Online Placement Test: The meaning of OOPT scores*. Oxford, England: Oxford University Press. Retrieved from [www.oxfordenglishtesting.com](http://www.oxfordenglishtesting.com)
- Puchta, H. (1999, March 28–April 1). *Beyond materials, techniques and linguistic analysis: The role of motivation, beliefs and identity*. Plenary session at the 33<sup>rd</sup> International IATEFL Annual Conference, Edinburgh, Scotland.
- Punch, K. (2005). *Introduction to social research: Quantitative and qualitative approaches*. London, England: Sage.
- Punch, K. (2012). *Introduction to social research* (2<sup>nd</sup> ed.). London, England: SAGE Publications.
- Punch, K. F. (2009). *Introduction to research methods in education*. Thousand Oaks, CA: Sage.
- Punch, K. J. (2011). *Introduction to social research: Quantitative and qualitative approaches* (2<sup>nd</sup> ed.). London, England: Sage.
- Purpura, J. (2007). *The Oxford Online Placement Test: What does it measure and how?* Oxford, England: Oxford University Press. Retrieved from [www.oxfordenglishtesting.com](http://www.oxfordenglishtesting.com)
- Rabab'ah, G. (2002). Communication problems facing Arab learners of English. *Journal of Language and Learning*, 3, 180–197.
- Rahman, M., & Alhaisoni, E. (2013). Teaching English in Saudi Arabia: Prospects and challenges. *Academic Research International*, 4(1), 112–118.
- Rao, R. (2014). Integration of Bloom's taxonomy and multiple intelligences in English classroom. *THE ELT PRACTITIONER*, 1(1), 1–3.
- Rauf, M. (2015). Best practices in English language testing at the university preparatory year programs. In A. K. Hamdan (Ed.), *Teaching and learning in Saudi Arabia*. Rotterdam, Netherlands: Sense Publishers.
- Reid, M. E., Clunies-Ross, L. R., Goacher, B., & Vile, D. (1982). *Mixed ability teaching: Problems and possibilities*. Windsor, Canada: NFER-Nelson.
- Reyna, C. (2000). Lazy, dumb, or industrious: When stereotypes convey attribution information in the classroom. *Educational Psychology Review*, 12(1), 85–110.
- Riley, T. (2015). "I Know I'm Generalizing but...": How teachers' perceptions influence ESL learner placement. *TESOL Quarterly*, 49(4), 659–680.

- Risager, K. (2006) *Language and culture: Global flows and local complexity*. Clevedon, UK: Multilingual Matters.
- Ritchie, J., Lewis, J., Elam, G., Tennant, R., & Rahim, N. (2014). Designing and selecting samples. In J. Ritchie, J. Lewis, C. M. Nicholls, & R. Ormston (Eds.), *Qualitative research practice: A guide for social science students and researchers* (pp. 111-145). Los Angeles, CA: SAGE.
- Robertson, R. (1995). Glocalization: Time–space and homogeneity–heterogeneity. In: M. Featherstone, S. Lash and R. Robertson (Eds.) *Global Modernities* (pp.25-54). London: Sage.
- Robson, C. (2011). *Real world research: A resource for social-scientists and practitioner-researchers* (3<sup>rd</sup> ed.). Oxford, England: Blackwell Publishing.
- Rossier, J. (2005). A Review of the Cross-Cultural Equivalence of Frequently used Personality Inventories. *International Journal for Educational and Vocational Guidance*. 5(2), 175-188.
- Rubib, H., & Rubin, I. (2012). *Qualitative interviewing: The art of hearing data*. London, England: SAGE.
- Ruecker, T. (2011). Improving the placement of L2 writers: The students' perspective. *WPA: Writing Program Administration*, 35(1), 91–117.
- Rumsey, M., Thiessen, J., Buchan, J., & Daly, J. (2016). The consequences of English language testing for international health professionals and students: An Australian case study. *International Journal of Nursing Studies*, 54, 95–103.
- Russell, M., & Haney, W. (1997). Testing writing on computers: An experiment comparing student performance on tests conducted via computer and via paper-and-pencil. *Education Policy Analysis Archives*, 5(3).
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55, 68–78.
- Sadler, D. (2009). Indeterminacy in the use of preset criteria for assessment and grading. *Assessment and Evaluation in Higher Education*, 34(2), 159–179.
- Sadler, I., & Reimann, N. (2018). Variation in the development of teachers' understandings of assessment and their assessment practices in higher education. *Higher Education Research & Development*, 37(1), 131–144.

- Sanford, K. (2005). Gendered literacy experiences: The effects of expectation and opportunity for boys' and girls' learning. *Journal of Adolescent & Adult Literacy*, 49(4), 302.
- Sawir, E. (2005). Language difficulties of international students in Australia: The effects of prior learning experience. *International Education Journal*, 6(5), 567–580.
- Schafer, W. D. (1993). Assessment literacy for teachers. *Theory into Practice*, 32(2), 118–126.
- Schönwetter, D., Clifton, R., & Perry, R. (2002). Content familiarity: Differential impact of effective teaching on students' achievement outcomes. *Research on Higher Education*, 6(43).
- Scott-Clayton, J. (2012). *Do high-stakes placement exams predict college success?* (CCRC Working Paper No. 41). New York, NY: Columbia University, Teachers College, Community College Research Center.
- Selvi, A. F., & Rudolph, N. (2018). *Conceptual shifts and contextualised practices in education for global interaction, issues and implications*. Singapore: Springer.
- Serrano, R., Llanes, A., & Tragant, E. (2011). Analyzing the effect of context of second language learning: Domestic intensive and semi-intensive courses vs. study abroad in Europe. *System*, 39, 133–143.
- Shay, S. (2005). The assessment of complex tasks: A double reading. *Studies in Higher Education*, 30(6), 663–679.
- Shepard, L. A. (1993). Evaluating test validity. *Review of Research in Education*, 19, 405–450.
- Shimizu, Y. (2002). Survey research on the use of placement tests at four-year universities in Japan. *Ritsumeikan Studies in Language and Culture*, 14(1), 231–243.
- Shin, S.-Y., & Lidster, R. (2017). Evaluating standard setting methods in an ESL placement testing context. *Language Testing*, 34, 357–381.
- Sikes, P. (2006). On dodgy ground? Problematics and ethics in educational research. *International Journal of Research and Method in Education*, 29, 105–117.
- Silverman, D. (2001). *Interpreting qualitative data, methods for analyzing talk, text and interaction* (2<sup>nd</sup> ed.). London, England: Sage Publication Inc.
- Simmel, G. (1950). *The sociology of Georg Simmel*. New York, NY: Free Press.

- Sinlapachai, R., Surasin, J., & Augkanurakbun, R. (2016). The use of online standardized tests as indicators of English proficiency according to the common European Framework of Reference (CEFR) at the English program of Chonkanyanukoon School, Chon Buri, Thailand. *HRD Journal*, 7(1), 63–72.
- Smith, T., & Strahan, D. (2004). Toward a prototype of expertise in teaching. *Journal of Teacher Education*, 55(4), 335–371.
- Smyth, A., & Holian, R. (2008). Credibility issues in research from within organisations. In P. Sikes & A. Potts (Eds.), *Researching education from the inside: Investigations from within* (pp. 33–48). London, England: Routledge.
- Spada, N. and Lightbown, P.M. (1989) Intensive ESL programs in Quebec primary schools. *TESL Canada Journal*, 7, 11-32.
- Stake, R. E. (2005). Qualitative case studies. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research* (3<sup>rd</sup> ed., pp. 443–466). Thousand Oaks, CA: Sage.
- Stiggins, R. J. (1991). Assessment literacy. *Phi Delta Kappan*, 72, 534–539.
- Stiggins, R. J. (1995). Assessment literacy for the 21st century. *Phi Delta Kappan*, 77(3), 238–245.
- Sturtz, A. J., & McCarroll, J. A. (1993, May). *Placement testing and student success: The first intervening variable*. Paper presented at the Annual Forum of the Association for Institutional Research, Chicago, Illinois.
- Sulaimani, A., & Elyas, T. (2018). A glocalized or globalized edition? Contextualizing gender representation in EFL textbooks in Saudi Arabia: A critical discourse analysis perspective. In A. Selvi, & N. Rudolph (Eds.), *Conceptual shifts and contextualized practices in education for glocal interaction*. Issues and Implications (pp. 55-76). Singapore, Springer.
- Susan, H., & Judith, I. (2005). Secondary school teachers' pedagogic practices when teaching mixed and structured ability classes. *Research Papers in Education*, 20(1), 3–24.
- Sweet, C. (2001). Designing and conducting virtual focus groups. *Qualitative Market Research: An International Journal*, 4(3), 130–135.
- Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4<sup>th</sup> ed.). Boston, MA: Allyn and Bacon.

- Taras, M. (2005). Assessment – summative and formative – some theoretical reflections. *British Journal of Educational Studies*, 53(4).
- Tashakkori, A., & Teddie, C. (2003). *Mixed methods research, alternative approaches and unresolved issues in conceptualization and design of mixed methods*. Thousand Oaks, CA: Sage Canada.
- Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches* (Vol. 46). Thousand Oaks, CA: Sage Publications.
- Taylor, L. (2006). The changing landscape of English: Implications for language assessment. *ELT Journal*, 60(1), 51–60.
- Taylor, S. J., & Bogdan, R. (1998). *Introduction to qualitative research methods: A guidebook and resource*. New York, NY: Wiley.
- Tenenbaum, H. R., & Ruck, M. D. (2007). Are teachers' expectations different for racial minority than for European American students? A meta-analysis. *Journal of Educational Psychology*, 99(2), 253–273.
- The College Board. (2003). *ACCUPLACER* ®. Retrieved from <https://www.collegeboard.org/>
- The College Board. (2009). *ACCUPLACER* ®. Retrieved from <https://www.collegeboard.org/>
- Tittle, C. K. (1989). Toward an educational psychology of assessment for teaching and learning: Theories, contexts, and validation arguments. *Educational Researcher*, 29(2), 149–162.
- Tolmie, A., Muijs, D., & McAteer, E. (2011). *Quantitative methods in educational and social research using SPSS*. Maidenhead, England: Open University Press.
- Tomlinson, B. (2012). Materials development for language learning and teaching. *Language Teaching*, 45(2), 143–179.
- Tomlinson, C. A. (2015). Teaching for excellence in academically diverse classrooms. *society*, 52(3), 203–209.
- Trifonas, P. P. (2009). Deconstructing research: Paradigms lost. *International Journal of Research and Method in Education*, 32(3), 297–308.
- Trigwell, K., & Prosser, M. (1996). Congruence between intention and strategy in university science teachers' approaches to teaching. *Higher Education*, 32, 77–87.

- Trigwell, K., Prosser, M., & Taylor, P. (1994). Qualitative differences in approaches to teaching first-year university science. *Higher Education*, 27, 75–84.
- Trigwell, K., Prosser, M., & Waterhouse, F. (1999). Relations between teachers' approaches to teaching and students' approaches to learning. *Higher Education*, 37, 57–70.
- Turner, D. W. (2010). Qualitative interview design: A practical guide for novice investigators. *The Qualitative Report*, 15(3), 754–760.
- Tymms, P. (2012). Questionnaires. In J. Arthur, M. Waring, R. Coe, & L. V. Hedges (Eds.), *Research and Methodologies*. (pp. 231–240). London: Sage.
- Tymms, P. (2013). Reviewing Rasch decisions. *Perspectives: Centre for Education Research and Practice* (February).
- Tymms, P. (2017). Purposes for educational research. In D. Wyse, N. Selwyn, E. Smith, & L. E. Suter (Eds.), *The BERA/SAGE handbook of educational research*. London: Sage Publications Ltd.
- United States Congress, Office of Technology Assessment. (1992, February). *After the Cold War: Living With Lower Defense Spending*, OTA-ITE-524. Washington, DC: U.S. Government Printing Office. Retrieved from [http://www.princeton.edu/~ota/disk1/1992/9202\\_n.html](http://www.princeton.edu/~ota/disk1/1992/9202_n.html)
- Unluer, S. (2012). Being an insider researcher while conducting case study research. *The Qualitative Report*, 17(Art. 58), 1–14. Retrieved from <http://www.nova.edu/ssss/QR/QR17/unluer.pdf>
- Ushioda, E. (1996). Developing a dynamic concept of motivation. In T. J. Hickey (Ed.), *Language, education and society in a changing world* (pp. 239–245). Clevedon, England: Multilingual Matters.
- Valiandes, S., & Neophytou, L. (2018). Teachers' professional development for differentiated instruction in mixed-ability classrooms: Investigating the impact of a development program on teachers' professional learning and on students' achievement. *Teacher Development*, 22(1), 123–138.
- Vandeyar, S., & Killen, R. (2007). Educators' conceptions of classroom assessment in post-apartheid South Africa. *South African Journal of Education*, 27(1), 101–115.
- Veenman, S. (1984). Perceived problems of beginner teachers. *Review of Educational Research*, 54, 143–178.



- Vogl, E., & Pekrun, R. (2016). Emotions that matter to achievement: Student feelings about assessment. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of human and social conditions in assessment* (pp. 111–128). New York, NY: Routledge.
- Wall, D. (1997). Impact and washback in language testing. In C. Clapham & D. Corson (Eds.), *Testing and assessment: The Kluwer encyclopedia of language in education* (Vol. 7). Holland: Kluwer Academic, 291–302.
- Wall, D., Clapham, C., & Alderson, J. C. (1994). Evaluating a placement test. *Language Testing*, 11(3), 321–44.
- Watanabe, Y. (2004). Teacher factors mediating washback. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 129–146). Mahwah, NJ: Lawrence Erlbaum.
- Weingartner, C. (1975). *Teaching as a subversive activity*. Harmondsworth, England: Penguin.
- Weir, C. (1990). *Understanding and developing language tests*. London, England: Prentice Hall.
- Weir, C. (2005). *Language testing and validation: An evidence-based approach*. London, England: Palgrave Macmillan.
- Weiss, D. J., & Schleisman, J. L. (1999). Adaptive testing. In G. N. Masters & J. P. Keeves (Eds.), *Advances in measurement in educational research and assessment*. New York, NY: Pergamon.
- Wells, C. S., & Wollack, J. A. (2003). *An instructor's guide to understanding test reliability, testing and evaluation services*. Madison: University of Wisconsin.
- Wesche, M., Paribakht, S. T., & Ready, D. (1996). A comparative study of four placement instruments. In M. Milanovic & N. Saville (Eds.), *Performance testing, cognition and assessment* (pp. 199–210). Cambridge, United Kingdom: Cambridge University Press.
- White, E. M. (1994). *Teaching and assessing writing* (2<sup>nd</sup> ed.). San Francisco, CA: Jossey Bass.
- White, J., & Turner, C. (2005). Comparing children's oral ability in two ESL programs. *The Canadian Modern Language Review*, 61(4), 491–517.
- Wiliam, D. (2001). Reliability, validity, and all that jazz. *Education*, 29, 17–21.

- Williams, M., & Burden, R. L. (1997). *Psychology for language teachers: A social constructivist approach*. Cambridge, United Kingdom: Cambridge University Press.
- Wilson, N., & Mclean, S. (1994). *Questionnaire design: A practical introduction*. Newtown Abbey, Co. Antrim: University of Ulster Press.
- Winter, G. (2000). A comparative discussion of the notion of 'validity' in qualitative and quantitative research. *The Qualitative Report*, 4(3), 1–14.
- Wiseman, A. W. (2010). The uses of evidence for educational policymaking: Global contexts and international trends. *Review of Research in Education*, 34(1), 1–24. <http://dx.doi.org/10.3102/0091732X09350472>
- Wolff, K. H. (1950). *The sociology of Georg Simmel*. New York, NY: Free Press.
- Wolming, S., & Wikstrom, C. (2010). The concept of validity in theory and practice. *Assessment in education: Principles, policy and practice*, 17(2), 117–132.
- Wood, R. (1991). *Assessment and testing: A survey of research commissioned by the University of Cambridge Local Examination Syndicate*. Cambridge, United Kingdom: Cambridge University Press.
- Woodrow, J. (1991). Locus of control and computer attitudes as determinants of the computer literacy of student teachers. *Computers and Education*, 16, 237–245.
- Wright, B. D., & Stone, M. H. (1979). *Best test design: Rasch measurement*. Chicago, IL: Mesa Press.
- Wyse, D., Selwyn, N., Smith, E., & Suter, L. E. (2017). *The BERA/SAGE handbook of educational research*. Los Angeles, CA: Sage Publications, Inc.
- Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, 58, 149–162.
- Yin, R. K. (1994). *Case study research: Design and methods* (2<sup>nd</sup> ed.). Newbury Park, CA: Sage Publications.
- Yin, R. K. (2003). *Case study research: Design and methods* (3<sup>rd</sup> ed.). Thousand Oaks, CA: Sage.
- Young, R. F. (2012). Social dimensions of language testing. In G. Fulcher & F. Davidson (Eds.), *The Routledge handbook of language testing* (pp. 178–193). New York, NY: Routledge.
- Zeidner, M. (1998). *Test anxiety – The state of the art*. New York, NY: Plenum Press.

Zeidner, M. (2014). Test anxiety. In P. Emmelkamp & T. Ehring (Eds.), *Wiley handbook of anxiety disorders* (pp. 581–595). Hoboken, NJ: Wiley.

## Appendices

### Appendix 1: ELI Level 1 Syllabus



## ELI 101 Course Syllabus

### Course Title

English Language Level 101

### Course Code

ELI 101

### Course Prerequisite

Oxford Online Placement Test (OOPT) score corresponding to Beginner proficiency level and below.

### Credits

There are NO credits for this course. Successful completion of ELI 101 is the prerequisite for taking ELI 102. Successful completion of ELI 102, ELI 103, and ELI 104 gives students the necessary credits to meet the Preparatory Year English requirement.

### Course Description

ELI 101 is a beginner course intended to provide students with a foundation from which they can advance from A1 Breakthrough to A2 Way-stage on the Common European Framework of Reference for Languages (CEFR). It is a seven-week module course with 18 hours of instruction each week.

### Course Goal

The course aims at helping learners to achieve an overall English language proficiency of beginner Basic User defined as A1 level on the Common European Framework of Reference for Languages (CEFR), developing “generative language use” to interact in a simple way and ask and answer simple questions.

### Course Objectives

The course is intended to accomplish its goal in one full academic module by developing students’ language skills to:

1. Read and understand basic expressions and short, simple texts.  
(READING)

2. Understand phrases and expressions related to areas of most immediate priority, provided speech is very slowly and clearly articulated. (LISTENING COMPREHENSION).
3. Engage in oral communications in very familiar situations, providing and obtaining essential information in simple mainly isolated phrases and sentences. (SPEAKING)
4. Write basic, simple sentences about people and places. (WRITING)
5. Demonstrate limited control of basic vocabulary and essential grammatical structures. (USE OF ENGLISH)

### **Main Course Textbook and Materials**

The core textbook is Adrian Doff, (2014), *A1 Starter Coursebook English Unlimited, Special Edition*, Cambridge University Press with the accompanying A1 Starter Self-study Pack.

### **Learning Outcomes (LOs)**

Learning Outcomes (LOs) are descriptions of what students will know or be able to do with the language as a result of instruction. The CEFR Tables found in the Teachers' Pack CD-ROM accompanying the teacher book lists the learning outcomes for this course in great detail. For convenience they are briefly summarized below.

#### **Speaking**

- can describe him/herself, what he/she does and where he/she lives
- can handle numbers, quantities, cost and time
- can ask and answer simple questions, initiate and respond to simple statements in areas of immediate need or on very familiar topics
- can ask and answer questions about themselves and other people, where they live, people they know, things they have
- can indicate time by such phrases as next week, last Friday, in November, three o'clock (Information exchange)

#### **Listening**

- can follow speech which is very slow and carefully articulated, with long pauses for him/her to assimilate meaning
- can understand instructions addressed carefully and slowly to him/her and follow short, simple directions (Listening to announcements and instructions)

#### **Reading**

- can understand very short, simple texts a single phrase at a time, picking up familiar names, words and basic phrases and rereading as required
- can get an idea of the content of simpler informational material and short simple descriptions, especially if there is visual support

## Writing

- can write simple phrases and sentences about themselves and imaginary people, where they live and what they do
- can write simple sentences at the CEFR A1 level about places and present and past events and actions

## Use of English Grammar and Vocabulary

- can demonstrate understanding of and use a limited range of vocabulary at the CEFR A1 level, including vocabulary for numbers, the time, days and dates, family, jobs, places, food and drink, countries and nationalities, common objects, clothes, colors, past time, future time and place expressions, transport and life events
- can demonstrate understanding of and use a limited range of grammar at the CEFR A1 level, including the verb 'to be' present and past tenses, pronouns and possessive adjectives, there's /there are/ this/these, singular and plural nouns, present and past simple positive, negative and question forms of regular and irregular verbs, can/can't, present progressive, past time expressions and frequency adverbs

## Assessment Overview

Student achievement is measured from a variety of assessment perspectives, including:

<i><b>Assessment Component</b></i>	<i><b>Units Assessed</b></i>	<i><b>Percent of Total Grade</b></i>
<b>Grammar and Vocabulary Quiz 1</b>	1 – 3	5%
<b>Writing Portfolio Task 1</b>	Mentioned with the task	5%
<b>Computer-Based Mid-Module Examination</b>	1 – 5	20%
<b>Writing Portfolio Task 2</b>	Mentioned with the task	5%
<b>Grammar and Vocabulary Quiz 2</b>	4-6	5%
<b>Writing Examination</b>	1 – 9	10%
<b>Speaking Examination</b>	1 – 10	10%
<b>Computer-Based Final Examination</b>	1 – 10	40%

## Description of Assessment Components

### Grammar and Vocabulary Quizzes

- Content: Each quiz has 10 Grammar and 10 Vocabulary questions from the units mentioned in the table above.
- Format: Grammar is a mix of fill-in-the-blank and multiple choice items. Vocabulary is all fill-in-the-blank from a word box which has more words than items.

**Computer-Based Mid-Module Examination**

- Content: Listening & Reading Comprehension, Usage of Grammar and Vocabulary
- Format: Multiple Choice Questions (MCQs)

**Computer-Based End-of-Module Examination**

- Content: Listening & Reading Comprehension, Usage of Grammar and Vocabulary
- Format: Multiple Choice Questions (MCQs)

**Speaking Examination**

- Content: Oral communication on a topic selected from the book, length and lexical variety appropriate to the course's CEFR level.
- Format: Answer 5 related questions individually and then participate in a guided conversation with a partner.

**Writing Examination**

- Content: Descriptive/Narrative text on a topic selected from the book, length and lexical variety appropriate to the course's CEFR level.

**Writing Portfolio Program**

- Details are provided by the classroom teacher each module, but the Writing Portfolio Program involves writing 2 drafts of 2 different writing prompts at specified times throughout the module.

## Appendix 2: Declaration of informed consent

### Declaration of Informed Consent

I agree to participate in this study, the purpose of which is to look at the validity of the OOPT test and if it predicts students' success in the English modules and later at their future major.

- I have read the participant information sheet and understand the information provided.
- I have been informed that I may decline to answer any questions or withdraw from the study without penalty of any kind.
- I have been informed that all of my responses will be kept confidential and secure, and that I will not be identified in any report or other publication resulting from this research.

I have been informed that the investigator will answer any questions regarding the study and its procedures.

Heba Nasem, School of Education, Durham University can be contacted via email: [h.m.nasem@dur.ac.uk](mailto:h.m.nasem@dur.ac.uk)

- I will be provided with a copy of this form for my records.

Any concerns about this study should be addressed to the Ethics Sub-Committee of the School of Education, Durham University via email (Sheena Smith, School of Education, tel. (0191) 334 8403, e-mail: [Sheena.Smith@Durham.ac.uk](mailto:Sheena.Smith@Durham.ac.uk)).

---

Date	Participant Name (please print)	Participant Signature
------	---------------------------------	-----------------------

I certify that I have presented the above information to the participant and secured his or her consent.

---

Date	Signature of Investigator
------	---------------------------



### **Appendix 3: Interview schedule**

Recommendation for the future, I want to know if they agree that we can make our own placement test that is suited to our students' culture, needs and abilities.

#### **Interview schedule:**

The questions will be divided into four categories: the first category will be a little about their background experiences. The second category will focus on the students, the third category will be regarding testing and OOPT. Finally, the last few questions will be on the modular system of the English course at the ELI. But before asking these questions I will start with a brief description of my study.

#### **1. Background information:**

- Could you tell me something about yourself?
- How long have you been teaching?
- How long have you been teaching in this institution?
- What level are you teaching now?
- Have you taught all four levels before?
- Could you tell me about tests that you have done?
- Have you designed any test before?
- Here is part of the OOPT test, what do you think about it?

#### **Rationale behind interview questions:**

Teachers' background will help reveals their actual teaching experience in general.

#### **2. The Students:**

- Which level do you believe has the most trouble with the placement test? Why do you think that is?
- Why do you think teachers are complaining about learners in lower levels having below average proficiency?
- Why do you think that? To what extent do you think that students' **financial background** reflects their success or failure in their English module?

- Why do think that students who were taught in **private school** are better in English than those who were taught in public school?
- Do you believe a learner's **language background** can affect a student's success or failure in their placement and throughout the English modules?
- How do you consider the orientation that is given to students to be important prior the test?
- Tell me any problems you might have faced in learners' using computers?

All these factors in **Bold** contributes to students' failure or success in the OOPT and the English modules.

All students are obliged to attend an orientation in the induction week to know more about the OOPT. These questions will help in revealing students' familiarity with the test and using computers and if that preparation is beneficial or not.

- Is there a difference in students' competence among **art** and **science** domains?
- Why do you think that science students are better?
- What other factors do you think that might affect their English competence?

### **3. The OOPT:**

- Can you tell me about your familiarity of the OOPT?
- Do you consider the OOPT to be an effective tool? How so?
- Do you believe that the OOPT results present an accurate representation of the learners' language abilities? How so?
- Have you come across learners who were misplaced in the wrong language level? Why do you believe that happened?
- What alternatives would you present to those learners to help remedy that situation?
- How **fair** do you think **the OOPT** is to **Saudi students**?

These questions (OOPT) are the focus of the study, through them I can know about teachers' perceptions and views on the validity of the OOPT. Also, it will strengthen and supports the predictive validity results that will be done quantitatively.

Through **fairness** questions, I want to know what is considered as fair in their perceptions, also if they know or heard anything about the content of the test, if there any item bias toward Saudi's culture in the test. Another main issue, is that the books for this year is Cambridge but the test is still Oxford. I want to know is it fair to use different test? Or it doesn't affect in any matte

Some teachers have mentioned before that the OOPT is **not fair** because it does not test speaking and writing. I want to know their opinions on that matter.

- What impact do you believe does having one curriculum publisher and a different OOPT publisher have?
- Do you believe the curriculum publisher and the OOPT publisher should be the same? Why?
- Would you say that **speaking** and **writing** should be included in the placement test?
- Why do you think that?
- How important are these two skills in a test?
- What are some methods can you suggest to help build our own placement test?

#### **4.The English module system:**

- To what extent do you believe that the English four modules will facilitate their future major?
- If students want to specialise in English Literature major, do you believe that the English modules have enhanced their English competence?

## Appendix 4: OOPT sample questions

oxfordenglishtesting.com/default/marticle.aspx?id=3074

This item tests a learners' understanding of literal meaning. Here the statement "Can I have an apple?" has the same literal meaning as "I would like an apple."

**Testing knowledge of similar meaning at A0 level**

**Task 2 for A0 level**

Testing a student in a short communication. There are three paragraphs of meaning. For the beginner test take the Placement Test Team at Oxford University Press.

**Example:** Test (pre-beginner)

Click to expand

Assign online homework today! Open an account with free sample practice tests.

Register for an **LMS account** and get started with **Online workbooks**

© Oxford University Press

OK

1 What does the boy mean?

☐ A I can eat an apple.

☐ B I have got an apple.

☐ C I would like an apple.

Read the conversation. Then, select the correct answer from the options below.

Boy: Can I have an apple?  
Woman: Yes, you can.

© Oxford University Press

Next >

oxfordenglishtesting.com/default/marticle.aspx?id=3074

meaning can be determined from the words in the sentence, to very implicit (also referred to as "implied"), where the meaning of the exchange is not stated directly. Being able to understand the meaning of an exchange, but also their context, is a key skill in a language.

**Example:** Test (beginner)

Click to expand

This item tests a learner's understanding of literal meaning. Here the statement "What's Ben like?" has the same literal meaning as "How does Ben look?"

**Testing knowledge of literal meaning**

**Task 2 for A0 level**

Testing a student in a short communication. There are three paragraphs of meaning. For the beginner test take the Placement Test Team at Oxford University Press.

**Example:** Test (pre-beginner)

Click to expand

Assign online homework today! Open an account with free sample practice tests.

Register for an **LMS account** and get started with **Online workbooks**

© Oxford University Press

OK

2 What does the man mean?

☐ A What does Ben like?

☐ B What would Ben like?

☐ C What do you think of Ben?

Read the dialogue. Then, select the correct answer from the options below.

Man: What's Ben like?  
Woman: He's nice.

© Oxford University Press

Next >

oxfordenglishtesting.com/defaultmarticle.aspx?id=3074

Use of English consists of three task types.

**Task 1 – Testing knowledge of grammatical forms**

Task 1 is designed to test your knowledge of grammar. In this task, you will be given a dialogue with one or more gaps. You will have to choose the correct word or phrase to complete the dialogue.

**Example:** Test your knowledge of grammatical forms at the A1 level (elementary).

**Task 2 – Testing knowledge of literal meaning**

Task 2 is designed to test your knowledge of the meaning of words and phrases. You will be given a short text and a list of words or phrases. You will have to choose the correct word or phrase to complete the text.

**Example:** Test your knowledge of literal meaning at the A1 level (beginner).

The test has been pretested and validated by thousands of students in more than 60 countries.

"I regard the Oxford Online Placement Test as one of the most reliable and quick methods of assessing a student's English level."

Dave Brown, Director of Studies, British Council Training Centre, Paris

enl Test  
jective  
s". Oliver  
dies, Key

12:23  
21/10/2016

**OXFORD** UNIVERSITY PRESS


**OXFORD** Online Placement Test

Online Placement Test

Use of English • Section 1: Approximately 30 Questions

Select a word or phrase to complete the conversation shown below.

6



Man: Who's that in the sports car?

Woman: It  be Frank. He hasn't got a car like that.

© Oxford University Press

Next >

Ok

## Appendix 5: Information sheet emailed to interview participants (Stage 2 interviews)

The interview questions will be divided into four categories: the first category will be about your background experience. The second category will focus on the students, the third category will be regarding testing and Oxford Online Placement test (OOPT). Finally, the last few questions will be on the modular system of the English course at the ELI. Below is a brief description of the questions that will guide the interview.

### **Main Questions:**

1. Which level do you believe has the most trouble with the placement test? Why do you think that is?
2. To what extent do you think that students' **financial backgrounds** reflects their success or failure in their English module?
3. Why do think that students who were taught in **private school** are better in English than those who were taught in public school?
4. Do you believe a learner's **language background** can reflect student's success or failure in their placement and throughout the English modules?
5. Is there a difference in students' competence among **art** and **science** domains?
6. Do you consider the OOPT to be an effective tool? How so?
7. Do you believe that the OOPT results present an accurate representation of the learners' language abilities? How so?
8. Have you come across learners who were misplaced in the wrong language level? Why do you believe that happened?
9. How **fair** do you think **the OOPT** is to **Saudi students**?
10. What impact, do you believe does having curriculum publisher and a different OOPT publisher have?
11. Would you say that **speaking** and **writing** should be included in the placement test?
12. What are some methods can you suggest to help build our own placement test?

## Appendix 6: Questionnaire

Dear Colleagues,

I would be grateful if you could participate in the following survey. This questionnaire is part of my research project in the PhD programme. The proposed questionnaire aims to look in depth at teachers' beliefs and their perceptions on the Oxford Online Placement test (OOPT).

The results of this survey will be used for the research purposes only. The responses will be handled with absolute confidentiality. It takes only few minutes to fill out the questionnaire. Thank you for supporting this study.

- Please tick the appropriate answer for each data item:

1. Age: ☐25-30 ☐31-36 ☐37-42 ☐43-48 ☐over 49

2. Qualification: ☐Bachelors ☐Masters ☐PhD

3. Country of qualification obtained:

a. Bachelors .....

b. Masters .....

c. PhD .....

4. Nationality: ☐Saudi ☐other .....

- Please choose the most appropriate answer:

5. What is your job role at ELI?

a. Full time teacher

b. Half load teacher with administrative work

c. Reduced load teacher with administrative work

If you have any other administrative work .....

6. How long have you been teaching?

a. 1 year to 5 years

b. 6 years to 11 years

c. 12 years to 17 years

d. 18 years to 23 years

e. 24 years and over

7. Do you know the content of the Oxford Online placement test?

- a. Yes                      b. No

8. If you answered the previous question with yes, do you think the content was fair for Saudi students? Cultural wise

- a. Yes                      b. No  
and Why? .....

*Please choose the box that best applies to you*

9. I believe that Oxford Online placement test is an effective tool, it places students to their appropriate level.

Strongly Disagree	Disagree	Do not know	Agree	Strongly Agree
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

10. I believe that Oxford Online Placement Test reflects student's abilities in the English Language.

Strongly Disagree	Disagree	Do not know	Agree	Strongly Agree
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

11. I think speaking and writing should be included in the Oxford Online Placement Test.

Strongly Disagree	Disagree	Do not know	Agree	Strongly Agree
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

12. I believe that students who were placed in level one struggle in the English module.

Strongly Disagree	Disagree	Do not know	Agree	Strongly Agree
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

13. I believe that students who were placed in level two struggle in the English module.

Strongly Disagree	Disagree	Do not know	Agree	Strongly Agree
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

14. I believe that students who were placed in level three or level four succeed in the English module.

Strongly Disagree	Disagree	Do not know	Agree	Strongly Agree
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

15. Students from private schools perform better in English than those in governmental schools.

Strongly Disagree	Disagree	Do not know	Agree	Strongly Agree
-------------------	----------	-------------	-------	----------------



☐ ☐ ☐ ☐ ☐

16. Students need to be prepared for this kind of test (Online test).

Strongly Disagree   Disagree   Do not know   Agree   Strongly Agree

☐ ☐ ☐ ☐ ☐

17. Science students perform better than Art students.

Strongly Disagree   Disagree   Do not know   Agree   Strongly Agree

☐ ☐ ☐ ☐ ☐

18. Student's financial background reflects their success or failure in the English module.

Strongly Disagree   Disagree   Do not know   Agree   Strongly Agree

☐ ☐ ☐ ☐ ☐

19. Traveling abroad and going to private schools enhance students' language competence.

Strongly Disagree   Disagree   Do not know   Agree   Strongly Agree

☐ ☐ ☐ ☐ ☐

20. Student's success or failure depends on other factors such as personal ability.

Strongly Disagree   Disagree   Do not know   Agree   Strongly Agree

☐ ☐ ☐ ☐ ☐

21. Language background considered being a major factor in student's success.

Strongly Disagree   Disagree   Do not know   Agree   Strongly Agree

☐ ☐ ☐ ☐ ☐

22. It is better to learn the language from an early age.

Strongly Disagree   Disagree   Do not know   Agree   Strongly Agree

☐ ☐ ☐ ☐ ☐

23. Motivation plays a vital role in student's success.

Strongly Disagree   Disagree   Do not know   Agree   Strongly Agree

☐ ☐ ☐ ☐ ☐

24. The English programme will positively facilitate student's future career.

Strongly Disagree   Disagree   Do not know   Agree   Strongly Agree

☐ ☐ ☐ ☐ ☐

25. I think that the modular system is successful.

Strongly Disagree   Disagree   Do not know   Agree   Strongly Agree

☐ ☐ ☐ ☐ ☐

26. Having mixed students (art, science) in one classroom are problematic.

Strongly Disagree    Disagree    Do not know    Agree    Strongly Agree

☐ ☐ ☐ ☐ ☐

27. The Oxford Online Placement test (OOPT) doesn't reflect everything in the student's competence in English proficiency.

Strongly Disagree    Disagree    Do not know    Agree    Strongly Agree

☐ ☐ ☐ ☐ ☐

28. If you have any additional point you want to add please write it down

.....

- I freely and voluntarily consent to be a participant in the research project on the topic mentioned above to be conducted by Mrs. Heba Nasem.
- I have been told that my responses will be kept strictly confidential. I also understand that my participation in this study is completely voluntary, and I may withdraw from it at any time without negative consequences. In addition, I am not obliged to discuss issues or answer questions that I do not wish to address or discuss. I will not be identified or identifiable in any report subsequently produced by the researcher.
- I have read and understand the above and consent to participate in this study. Moreover, I understand that I will be able to keep a copy of the informed consent form for my records.
- I am willing to give my consent for you to use the information that I have provided in this questionnaire in your research project

.....

.....

Participant's signature

Date

## **Appendix 7: Sample of a transcribed interview**

Interview with Stage 1 Participant 11

Me: Could you tell me something about yourself? What is your background?

Participant 11: I have a bachelor in English literature and a master in English linguistics, I have been teaching for almost 10 years now, I do have some educational experience but mostly my experience in linguistics.

Me: What level are you teaching now?

Participant 11: level 101

Me: Have you taught all levels before?

Participant 11: I have taught all levels

Me: Can you give an idea about your opinion about tests in general? Do you like tests, does students like tests?

Participant 11: I think am old school I do believe in tests, I do believe that if they were written probably they evaluate the students, it is not only about marks it tells students herself where are her weaknesses and her strengths and they also help the teacher to elaborate on her method like if all the class did not do well in the vocabulary test then she maybe she should concentrate on teaching vocabulary in the classroom, so I think it is a good evaluation method for both teacher and student.

Me: have you designed any test before?

Participant 11: yes we used to write our own tests before in the Eli, I did write couple of tests.

Me: From what you have seen (sample of OOPT) what do you think about it? Is it suitable for our students? Are the instructions clear?

Participant 11: I think the test is really clear, it is very suitable for student, I just have a comment about the test that some part are not culturally related like some saying or some concept do not relate to our culture so our student wont know how to answer it which will affect their evaluation.

Me: which level do you believe has the most trouble with placement test? Why do you think that?

Participant 11: I think the test places student beneficially in levels 2,3, and 4. But the problem with level 1 is that all students who are less than level 2 they are placed in level 1, we have a range of students who are suitable to level 1, and we do have students who do not have even the basic of English, we can see this clearly when the students move to level 2 some of them move with an A and some of them fail and some of them move with D, those students when they develop throughout the levels they start struggling in level 3 and 4 and start repeating them two or three times, because you feel that they were not placed correctly in level 1, maybe they needed like preparation course. We don't usually have any trouble with students who were placed in levels 2,3 and 4 they usually are good. For example, if a student were placed in level 2 and when she moves to level 3 she benefits greatly and develop really nicely in terms of language skills. Even students who were accepted for level

3 and 4 but students who were accepted on level 1, let's say 20% of them are not suitable for that level.

Me: so what do you suggest? Do they need more time in each level?

Participant 11: maybe they should be tutored or taught in a different way, for example, if we are placing students who scored less than ten in level 1, so from 5 to 10 they should be placed separately and from 0 to 5 should be placed separately. So level 1 should be treated this way.

Me: why do you think teachers are complaining about learners in lower levels having below average proficiency? 7:35

Participant 11: we do have a problem when it comes to level, we see a gap between level 2 and 3, we see students pass 1 and 2 easily, and then when they start level 3 they start struggling it is a big jump in terms of grammar, writing, listening, speaking between these two levels, and students let's say they take it easy in level 1 and 2 because they are really easy, and we keep on telling them it is going to get harder but they don't act until they reach level 3 and you see a great drop in marks between level 2 and 3. The other problem as we discussed that they needed more attention in level 1 they need more education.

Me: To what extent do you think that students' financial background reflects their success or failure in their English module? Private school

Participant 11: if we are talking about students who are financially supported they go to private school and we see a great difference students who were placed mostly in level 3 and 4 and some times 2 and they have the basic of language skills, so they do perform much better than other students who were studied in public schools. The main reason is in private school they start teaching English in grade 1 (6 years old) and in public school they start from grade 7, which is a huge difference between grade 1 and grade 7.

Me: Do you believe a learner's language background can affect a student's success or failure in their placement and throughout the English module? 10:36

Participant 11: of course if she has background in language skills she will develop more beneficially in the four levels, but if her background is really basic of course she won't benefit as much. Of course it is a big difference, for example student who have travelled before and were exposed to the language and who have sometime when they travel because of course we are going to the financial background is she knows that she will be using the language every summer and if she like open to American movies, American shows, she will have motives to study the language, some times we deal with students who never travelled outside Saudi Arabia, she never needed the language and she never even watch American or English programme she would be less motivated, she will only has educational purposes. Another motivation thing is that what does the student want to specialise in, what is her aim, now if she wants to study medicine or any thing related to science she knows that she needs the English, but if she is specializing in Arabic, geography or history, she knows that English is for her is like a waste of effort and that makes a big difference in our students.

Me: How do you consider the orientation that is given to students to be important prior the test?

Participant 11: it is very important and very beneficial, it is very educational. We deal with students as if it is their first electronic exam. So we try to show them how to move between

pages and how to choose the answers so it is basically teaching the skills of taking the exam it is not related to the language itself. It is more technical.

Me: Tell me any problems you might have faced in learners' using computers? 13:37

Participant 11: yes of course especially in the placement test, some students they do not even know how to log in they do not know how to put their password and user name and some time we do this for them, we do face students who don't even have a computer at home, these students come from the country side. One of the student told me my parents refuse to install any program in the computer, not even word program. Another student told me that her parents refuse to participate in the English forum. So we do face these issues we always have technician in the rooms, we show and help the student with any technical problem. We guide her through the exam and till we make sure that she logged out correctly. So we don't waste her exam.

Me: Is there a difference in students' competence among art and science domains?

Participant 11: of course there is a big difference; usually science students are much better in language than art, maybe because they are motivated or they want to specialise in major that do need English language but mostly mostly mostly I can not remember when did we compare an art group with a science group and we found art group is doing better in language. So mostly science group are better. We believe that who ever get high GPA in high school they go to science domain, and students who get low GPA they go directly to art even if they want to specialise in science. Art students they are not as dedicated or motivated and smart as science students. Also, the art students usually lazy and they do not want work hard.

Me: what other factors do you think that might affect their English competence?

Participant 11: we talked about motivation it is very important factor, the second factor is most of the time the students are really embarrassed of their language skill. She is embarrassed to make a mistake in front of her teacher or her friends, because we do have this cultural thing to make fun of mistakes like we do make fun when some one pronounce the wrong word, so we do have this in the classroom it is a cultural thing even if some one speaks Arabic and he makes a mistake even if he a native we make fun of that. I think that affect their performance in classroom especially when it comes to speaking and that is 80% of language we need the language to speak and to communicate. Even when we put them into groups because we are dealing with the same nationalities, for example when you study English in the States or Britain, our friends the group are not Arab, that a problem we face as teacher once we put them into groups they speak in Arabic, they only write the answer in English, so we lose 80% of the benefit of putting them into groups. So that is one of the struggles we face in the classroom. Lets say motivation and embarrassment the second reason and there is that some time we face this group of students who lost hope in English she is like no I will never will be able to speak English, I can not do it. And they believe it is like any other subject like I will read at home I will memorise and I will speak English, I always tell them the language is a skill it is not only about knowledge, I compare it to driving to swimming if you memorised a book about driving you wont be able to drive until you practice that is exactly the same in language you have to practice writing, reading communicating. Some times students are convinced many times they already hope before even they start university they lost hope since they were at high school or elementary. Sometime teachers who taught them in school they were not that good and they used to

memorised to pass the exam. Some of them have that mentality or attitude toward the English language that who needs that language after all! They won't need it where they come from. We told the test committee that please we don't want in writing exam that kind of questions that ask students where did travel abroad? You can say where did you travel inside Saudi Arabia? Because we are dealing with a government university, rarely we find girls who have good financial background and have traveled abroad. Those students we usually find them placed in level 1.

Me: Can you tell me about your familiarity of the OOPT?

Participant 11: actually I have taken it once, I wanted to apply for something and they want me to take it and I was so curious while taking it so I was observing the exam more than taking it. In general I like the test I think it is a good evaluation tool.

Me: Do you consider it to be an effective tool? 24:32

Participant 11: yes definitely, as a teacher I have seen the exam and I see the students every module, so I see how they are placed and I believe they have placed beneficially as I said 2,3,4 placed perfectly, the only problem is with students who were placed in the first level.

Me: Do you believe that the OOPT results present an accurate representation of the learners' language abilities? How so?

Participant 11: well actually no, it is a general exam so when we put students into level, we are talking 1,2,3, 4 so the students who were really good in vocabulary and reading they get really high score, the exam has no writing part and this is really... what mostly clear is writing part. Of course we do face students who she is basically is a good student or she knows how to apply the rules, she knows how to correct the right answer when it comes to grammar, vocabulary and memorise a lot of words let's say in level 2 or 3 but she does not know to write a paragraph. Speaking is a problem in all levels I don't think even if we evaluate them we can do nothing about them like we are in this country they are not going abroad we are not using the language outside the classroom we do not communicate or use it all the time until she finishes her first year at university, when she moves to second year if her major needs English she will use it and practice it. One of the major problems we face is writing definitely.

Me: Have you come across learners who were misplaced in the wrong language level? Why do you believe that happened?

Participant 11: I always have my own opinion regarding that matter, they placed correctly but moving through the levels we may find some struggle, because some times if she is placed in level 1 I don't know if it is about her performance or her attending classes or her teacher, when she move to level 2 we do not see her as a level 2 student, so just to make the answer clear, the first module after the placement test most students are placed in the right level, we are not talking about exception or students who guessed for the exam we are not talking about exception, we are talking about the majority the first module we see them placed correctly but then moving thorough out the levels we see some struggle maybe because the modular system which is they spend only seven weeks. The modular system is really not working, you get students to know you and they start to talk to you and encouraged to speak and then they move and another teacher will teach them. They usually take two weeks to get used to the new teacher.

Me: have you ever faced in a classroom either good students or very weak in the same classroom?

Participant 11: no, what I usually face is that students who were placed in level 1 lets say and when she moves to level 3 I find her very weak. Sometimes student do not do really well in the placement and when we evaluate her as teachers we can see that she is better than what she were placed in so we ask to give her another placement test. She might be very sick or has family issues. And also there are some students they did not take the test at the first place, they don't want to take the test. They might be out of the city, some of them we find them very good in English.

Me: how fair do you think the OOPT is to Saudi students? From what you have seen do you think that there is any item biased toward Saudis? Or the questions form is it different that what they have used to have

Participant 11: as I mentioned before there is a cultural part, we are talking about lets say foreign metaphor, saying, and I can assure you that no students will know about, that part should be changed, we should have special edition for Saudi students. And when I say special edition, I don't mean to change the names from George to Abdullah!! No, this is not a special edition we are not talking about names. We are talking about the culture itself, am talking about subjects like skiing we don't ski we don't have snow or rugby sport we don't play rugby, our girls they don't even play sports. I used the dictionary when I taught about rugby for the first time. Our girls don't play outdoor sports. It is very rare we do not play football, or basketball, it is culturally oriented country. We don't even have places to play these sports, even when we put them to gyms the last they can do is aerobic and zumba dancing classes if they are allowed. So when you want to talk about a city for example don't talk about Vancouver, talk about a city that a bit similar to our country like tour famous places. Don't give me a whole unit about sports, and when it comes to our students they need a unit about cooking things that they are familiar with it. I am not saying that we should not educate them but if I have the chance to teach them a hundred vocabulary words in the module, lets use a hundred vocabulary words that she will be using it. I always like the chapter or unit about restaurants, she may use them when she go to a restaurant and the waiter is not a Saudi, for example from Philippine she may use it, I always like unit about shopping, because we do have many shop assistance that do not speak Arabic we have people from Philippine, Sirilanka and other cities and they only speak English. So I don't like to spend a whole unit talking about sports that 80% are not played in Saudi Arabia because of the weather. So am talking about books and placement test. so making a special edition is not about putting scarfs on the head, or the man wears our own costume, no it is not what we are talking about.

Me: what impact do you believe does having one curriculum publisher and a different OOPT publisher have?

Participant 11: this only started last year, am sure it is affecting our students. Am sure it affected the placement of our students in the levels. When am talking about all the previous question I am talking about the last five years that we have been teaching Oxford, so I am not talking about the last year only when we changed the books, I think it is a major mistake, students may be placed correctly to the right level but I am sure if that exam was exactly like the book, the output would be much more specific.

Me: Do you believe that we have to build our own placement test? What are some methods can you suggest to help build our own placement test?

Participant 11: but we are not using our own books, I think we should always go with the same company of the book itself. Or may be it would be a good idea, because we are designing the test that is moving the student from level to another, so if we are able to evaluate the student and move her from level 1 to 2 and from 2 to 3 and so on, I think we would be able to do our placement exam. So this might be a good idea because we know our students better than any one else. We have already our own test bank, the OOPT is costing tons of money, we can have a committee who can spend a year on organizing and building the test, we have a specialised teachers in writing exams. We can have a trial on two group of students or ten groups. 40:40

Me: To what extent do you believe that the English four modules will facilitate their future major?

Participant 11: I don't think it is beneficial at all, for many reasons lets say: one is too short, please don't compare our program to a program that is taught in Denver Colorado while student is exposed to language for 24 hours. The students study for seven weeks and it is not eight it is seven with exams and every thing. Once the student is comfortable in the level with the teacher and with a group of students, she is moved to another level with other teachers with other students, and that is like taking her five steps backward that is one. Two, the whole university system is a semester based system, all other subjects taught in a semester based system, only the English is a modular system so we are interfering with everything else with other classes...

Me: remind me why are we following the modular system? Is it for accreditation?

Participant 11: no, we caught the accreditation after we started the modular system so we cannot change it back because we will be changing a major thing in accreditation. It was a one-man decision he believed in the modular system and he wanted to move that to the whole university for the first year (foundation year) to be a modular system, other department refused and we are the only department who agreed and started it, after that like a year or two we got accredited with the modular system. It is one of there conditions that once you get accredited you cannot change anything. If we made a major change we have to get evaluated again and that is huge procedure.

Me: If students want to specialise in English Literature major, do you believe that English module have enhanced their English competence?

Participant 11: no, the same reasons the English module does not give enough for the students, they need more time so it can affect her language skills in general whether she specialises in English or any other major.



## **Appendix 8: Sample of initial coding document**

Results:

### **1. Teacher's view on OOPT:**

The interviews have shown that teachers seem to be satisfied with the test and validate the Oxford Online Placement test in term of serving its purpose in placing students in their appropriate level. Nine of eleven teachers thought that the test was an effective tool:

Very valid exam with very valid result, this exam is very user friendly unlike other exams (Interviewee 01)

I really like it, we all are happy about it (Interviewee 05)

It is very beneficial because the outcome of the test is classroom that has students almost the same level (Interviewee 011)

### **2. Does OOPT reflect student's abilities?**

Five teachers have indicated one major weakness in OOPT; that is it does not tackle the four language skills:

It doesn't test all their skills, it doesn't test speaking or writing (Interviewee 06)

One teacher has emphasised the importance of testing writing in placement test:

Writing is really what shows you what confident and proficient student (Interviewee 05)

Another teacher reported that students in level two and level three were below their actual level and that OOPT did not reflect their ability especially in speaking and writing:

This can be explained by the fact that the OOPT doesn't focus on these certain skills (Interviewee 010)

One teacher added that the exam is 'meticulous' in term of reflecting their real ability:

It places them in the right level, but it doesn't show the ability of the student, it is not deep, it is so general but beneficial in only placing students in the right level and not in showing their ability (Interviewee 011)

Another teacher claimed that:

In what it claims to do I think it does, it is more or less reflective of what their skills are in receptive skills (Interviewee 03)

Four teachers think that Oxford Online Placement test reflects student's ability. One of the teachers added that:

I think that placement test does have some kind of connection, it measures student's ability may be 70 to 80 percent (Interviewee 09)

### **3. Placing students to their appropriate level:**

The Head of Exam unit claimed that each year students complain about their levels, some of them think they have to be in a higher level and others feel that they should be in a lower level. Accordingly, they have decided that teacher should evaluate her students in classroom and can move any student who she thinks that has to be in different level. The Head of Exam Unit has indicated that last year only 2% of students who have retake the placement test and changed their actual level according to teacher's evaluation.

Five teachers have added that for the majority it works very well with some exceptions. One teacher advised that teachers may find classes with multi-level students and teacher should look into students' differences and help student individually:

English classes are like any language classes, will always be the same. You will always get spiky profile of language skills in one classroom, learning needs are always different and learning needs are always individualise. Individual differences will always be there (Interviewee 02)

Eight teachers have contradicting opinions in regard to placing students to their appropriate levels. Three teachers think that placing students in level 1 and 2 is usually quite good. One of the teachers feels that:

It is pretty normal to find in every class few students who are over or below (Interviewee 05)

On the other side, one teacher faced problems with students that were placed in level 1 and 2. They had problems in grammar, vocabulary and speaking:

The reason behind this problem was guessing the answers and their English language is very weak (Interviewee 09)

A similar situation happened with one of the teachers where she faced a problem with one of the students in level 4, the student was very weak and could not understand anything in class:

She guessed the placement test that's why she has been in level 4 (Interviewee 08)

Another teacher experienced a problem with level 2:

I faced a 102 student who is better to be placed on 101 level. Her comprehension, writing and speaking skills were not of a 102 student (Interviewee 010)

Whereas, one teacher has stressed the effectiveness of the OOPT in placing students in their appropriate levels without having any problem with that but:

Other teachers have faced some problems, where students haven't pay attention to exam and guessed most answers (Interviewee 011)

**4. To what extent do you think that students who had high scores on the OOPT will be successful in their four modules?**

This question has definite answer according to many teachers, but one of them has described it as:

If students were placed in level 1 or 2 it means their English proficiency was not good, this is a proficiency test, achievement test, we are testing what they have studied (Interviewee 03)

What can be understood from this question particularly is that students who did well in the OOPT test and had high scores will definitely be on high levels (3 or 4). Since the OOPT originally should place students to six levels but KAU decided to teach till level 4. The Head of Exam Unit added that:

For good students, it's a piece of cake, she will pass it with excellence (Interviewee 01)

Similarly, another teacher have emphasised that:

If the student was placed in level 4 it means she is excellent, she is able to manage and deal with level 4 and she excels. If students were placed in level 1, you find them struggling (Interviewee 06)

Another teacher declared that:

Sometimes I think that students in level 4 they don't need my teaching because their English language skills are very very good (Interviewee 02)

### **5.OOPT Fairness:**

A few teachers (three teachers) commented that they haven't seen the content of the test in order to answer this question. Others have mentioned that it is not a fair test culturally wise, about four teachers have said that OOPT is not related to Saudi Culture unlike the books, it is a special edition, which is related culturally to Saudis. One teacher has added an important point regarding test fairness toward certain students who did studied in governmental school where English language has been exposed for the first time on the 7<sup>th</sup> grade. These students sometimes don't know the simplest rules in English nor can differentiate between the Alphabets:

The books in government school are basics and did not use or practice the language outside classroom (Interviewee 010)

This particular teacher suggested that if:

The OOPT has a zero level. Students who had a very limited exposure to English can be placed on this level

One teacher questioned the fairness of the test since it does not contain the four language skills and it does not reflect nor cover everything in the language:

I don't think it's fair to students, there are two major skills are not tested. Some students are good in speaking or writing but they don't get the chance to be tested (Interviewee 02)

Another teacher thought that question is really important and thinks that Saudi students are not prepared for that kind of test:

Students are not acquainted with these types of exams, they didn't have the right practice, they are not prepared. They have to be prepared and have the right skills to do good in the exam (Interviewee 09)

Lastly, one teacher (Interviewee 06) mentioned that some of the students don't use or own computers and this might affect their grades

**6. To what extent do you think that student's financial background reflects their success or failure in their English module?**

Most of the teachers about (eight teachers) answered this question with a strong agreement on the fact that financial background does affect their success in English language. Since the English language from my point of view considers as a prestigious language especially for Arabs learners. One of the teachers explained it as follows:

Students who can travel and go to restaurants and use the language are very anxious to learn the language. On the other hand, students who never travelled and never had the chance to use the language are not anxious they are not motivated and they don't feel the need of studying English (Faculty member 011).

Another teacher emphasised on travelling abroad and going to private school help students in a way that:

This gives them an opportunity to get exposed to the language in its purist form, learning will be enhanced and the language competence will be better (Faculty member 02)

Students who had better income can afford going to private schools where they can study the language from an early age with talented teachers and some times native speaker

teachers. Unlike public school that does not provide the English language till year 6 and teach very limited English.

Three teachers thought that financial background could not affect their success or failure rather it depends on many factors such as language attitude and language ability. One teacher explained:

I cannot generalise this because some students are excellent despite their financial background. It might be a personal ability they are quick learners and pick the language easily (Faculty member 010)

**7. To what extent do you think that student's language background reflects their success or failure in their English modules?**

All teachers agreed that language background is a major factor for student's success, where they can move easily from one level to another (Interviewee 011). But it should be acquainted with other factors:

It is one of the factors because if they have the English capabilities but they are not studying then it could affect, but as long as they have the English capability and they are hard worker students then both these factors working together will help in achieving (Interviewee 09)

Another teacher added that Language background is an important factor but it is better to learn the language from early age (Interviewee 02). Finally, one teacher claimed that language background is important but:

Again I believe it is a matter of different personal abilities, styles, and of course motivation (Interviewee 010)

**8. Do you think the English four modules will facilitate their future major?**

The results of this question are contradicting, five teachers think that the English modules will not facilitate their future mainly because it is too general and offers basic information about the language. One example of that:

Honestly no, the four module will help them pass the English level but without giving them the English capabilities needed to help them in business or finance or in

English language or any department, they need specified English or English for specific purposes in order to help them achieve that (Interviewee 09)

One of those teachers (Interviewee 01) criticised the books (Oxford books). She claimed that some of the books are a repetition of the previous books, some are too easy and the others are too difficult.

Two teachers (Interviewee 04 and 07) think that the English modules may help students in their future departments; if that department is English relevant and their books materials are in English.

On the other hand, four teachers think that the English modules will definitely help and facilitate them in the future wither in their study or in their future careers. One teacher declared that:

It depends on many things like: how good is the teacher, student's motivation and student's differences (Interviewee 02)

Another teacher believes that it depends on students, some students take the English module more seriously and those might succeed and achieve better than the other students who don't take it seriously (Interviewee 06)

Lastly, one teacher (Interviewee 05) considers that the English course as:

A preparation for upcoming assessment when pursuing their higher education. We teach speaking, writing, reading and listening, all that included in the IELTS and TOEFL exams.

**Additional points:**

A senior teacher (Interviewee 03) who has over 26 years of experiences criticised the modular system of the English course for the foundation year. She claims that the ELI (English Language Institute) is the only department at the university that follows modular system instead of semester system. She added that this creates a lot of conflict because:

There is not enough time to test the material, you try to stuff two books just too much, you can teach it but I don't think student can digest it (Interviewee 03)

Another teacher (Interviewee 010) has the same point of view as the previous teacher. In addition, this teacher mentioned that the modular system is not successful for many reasons and these might affect the student negatively:

Students in foundation year study many courses beside the English which make them busy and don't have time for all subjects, it demotivates them, they complain about the long hours before or after which they have other classes. Many of midterm exams happen during English classes, which force them not to attend the class (Interviewee 010)

A third teacher (Interviewee 07) has the same opinion regarding the modular system. She believes that the English course puts a lot of pressure on students because it is stuffed with many quizzes:

They think of it as more than a burden than a learning opportunity (Interviewee 07)

Only one teacher (Interviewee 05) believes that students might benefit more from the modular system, because during the modules they pass by a set of teachers and that is beneficial in her opinion, students might get one thing from every teacher.

The senior teacher (Interviewee 03) faced another problem, which is having mixed students in one classroom that means having Art and Science students in one classroom. She suggests that from the beginning of the year, teachers and students should know what they want to major so they can be divided into specialised classes. For example, Art classes and Science classes. This way teacher could tailor the materials on the curriculum to be in that field.

Two teachers think that Science students are better than Art students. One of the teachers (Interviewee 05) described these students as:



Science students work harder, they worry about grades and they do well in the exams, they take their education more seriously. Art students are weak in academic skills, but they have good social skills (Interviewee 05)

The second teacher (Interviewee 04) claims that Art students are problematic; they are weaker than Science students. She suggests:

From my experience I see that this curriculum is better to have it as levels. But for Science students I suggest to have a higher level that has more scientific vocabulary, and sections that teach them how to write a proposal (Interviewee 04)

Another problem teachers have mentioned is regarding the English course. One teacher (Interviewee 06) thinks that any student who chooses to study and specialise in English Literature should have to pass an acceptance test because:

They don't believe that the English course is not really good, is not really serving them, it is not high enough because to be honest it's very basic (Interviewee 06)

A senior teacher (Interviewee 02) suggests that the OOPT need to be looked at again, because it doesn't reflect everything in the student competence in English proficiency. And according to (Interviewee 01) all skills should be balanced in academic ground.

Lastly, one teacher (Interviewee 06) criticised student's learning style. She thinks that students don't want to work hard and they don't want to learn because they don't really know the importance of the language in their lives. She suggested that they have to work on themselves:

A lot of students not autonomous learners that's another problem and trying to get them to become autonomous learners is really challenging for teachers. It is hard because culturally it's not acceptable here. The teacher is every thing in class and suppose to spoon-feed students and that's another challenge (Interviewee 06)

## Appendix 9: Sample coding document

### Importance of Preparation for OOPT:

Teacher	Orientation	Using Computers
1	it is very important very vital for the students to be able to take the test	the fact that they exist that this person is exist until last year. It was very chocking, but we had to deal with it. I told her to learn how to use computer because every one uses technology, you wont be able to do anything in university.
2	it is important for students just to know how to go through the exam more faster but it would not reflect the result of the placement test	they all know how to click the keyboard and use the mouse.
3	if there is an orientation to show them how to answer that will be very useful and helpful, it will distress them and take the anxiety out. I think it is very good idea.	the generation is really computer generation but the only problem that we always face is that the internet connectivity, I think that is the only problem, otherwise the students are very good users of computers because they have smart phones and smart devices all the time
4	am with the orientation	
5	they needed like preparation course.  it is very important and very beneficial, it is very educational. We deal with students as if it is there first electronic exam. So we try to show them how to move between pages and how to choose the answers so it is basically teaching the skills of taking the exam it is not related to the language itself. It is more technical	some students they do not even know how to log in they do not know hoe to put their password and user name and some time we do this for them, we do face students who don't even have a computer at home, these students come from the country side. One of the student told me my parents refuse to install any program in the computer, not even word program
7	I have no ides about this orientation am sorry. But I would assume that it will be really positive because if all	this generation kids are really good in technology and so.

	<p>learners are taking for example the ielts the toefl exams if they have enough preparation or orientation before the test they will do better. If they take any sort of orientation any sort of practice if they are familiar with the format of the test, the number of the questions everything am sure it will have a positive affect especially because it is a computerised test and they are used to pen and paper test</p>	
8	<p>it is not that useful, it does not really explain the test, it is only tells them what to press to skip or to go to the other page and that set. Also, it tells them for example that the listening will hear it only once and not twice. It does not really explain what I want exactly.</p>	<p>all of students know how to use computers, but the technical problem that happens in the test is a problem. Each lab has around thirty students and only one or two technicians serves the floor and the timer of the test is working and students are waiting to get help. It should be one technician for every lab. This will definitely affects on students' results of the placement.</p>