University of Southampton Research Repository

UNIVERSITY OF SOUTHAMPTON

FACULTY OF SOCIAL SCIENCES

<u>SCHOOL OF EDUCATION</u>

The Predictive Validity of the Admission Standards in the College of Education at King Faisal University in Saudi Arabia

By

WALEED GHAZI ALSHAMMARI

Thesis for the degree of Doctor of Philosophy

November 2020

ABSTRACT

THESIS TITLE: The Predictive Validity of the Admission Standards in the College of Education at King Faisal University in Saudi Arabia

WALEED ALSHAMMARI

Universities face numerous challenges each year, including the process of making decisions concerning the admission or otherwise of applicants (Tesfa, 2013). This could be compounded by an increase in student numbers and a decrease in university resources, predicting future academic success (Alghamdi, 2007). Educators and admission officers together try to decide what leads to the success of learners at certain colleges or on particular majors within universities. Administrators responsible for the admission policy need to be accurate and objective when making such decisions, using suitable admission standards which help in reaching a decision characterised by equity, accuracy, and objectivity (Alsaif, 2005).

The purpose of the present study is to investigate the predictive validity of the current admission standards applied at the College of Education at KFU in Saudi Arabia and explore which score among the current criteria used offers the strongest contribution to students' academic success. Furthermore, since this study attempts to include students from both gender groups, and very few studies have included both genders at the general level, and, to the best of my knowledge, none has been done in the Saudi context, this study aims to explore any possible variation between the criteria items in terms of gender grouping. Additionally, since none of the previous studies have addressed the issue of students changing their major after initially being admitted to certain majors at the university, this study attempts to explore the academic performance of students who changed their major after starting their university study. The participants in this mixed methods research largely drew on two resources: the first resource refers to the data that was collected from the

Admission and Registration Office in the Education College. The database includes all full-time students (males and females) who have attended the Education College at King Faisal University from the academic year 2010 up until their graduation in 2014. The sample did not include students who had left the Education College at KFU before the end of the academic year 2014 and any students who had not graduated by the end of the academic year 2014. The total number of participating students was 693. In addition, the researcher conducted face-to-face interviews with 8 academics who work at the Education College in KFU and who teach a number of education courses. These lecturers and professors were interviewed about a range of experiences and practices. Results indicated that a statistically significant relationship exists between the student accumulative rate in High School (SGPA) and the accumulative rate in the College of Education (CGPA) at King Faisal University at Alahsa ($r = 0.562$, $p<0.01$), between the General Aptitude Test (GAT) and Education College GPA ($r = 0.324$, $p<0.01$), and between the Achievement Test (ACT) and Education College GPA ($r = 0.268$, $p<0.01$). High School GPA is the most important factor in predicting the performance of students in the Education College, followed by the Aptitude Test, then the Achievement Test. Beta coefficients were 0.512, 0.163 and 0.006 respectively. Regarding the result, it was clear that females exhibited better performance compared to males in both the General Aptitude Test and the High School percentage. In addition, the students who changed major had a higher High School percentage mean compared to those that did not change major, and the mean difference was statistically significant. On the other hand, for the General Aptitude Test, those that did not change major had higher mean scores compared to those that changed major, and the mean difference was statistically significant.

Table of Contents

Table of Tables

Academic Thesis: Declaration of Authorship

I, Waleed Ghazi Alshammari, declare that this thesis and the work presented in it are my own and this thesis has been generated by me as the result of my own original research.

Title: The Predictive Validity of the Admission Standards in the College of Education at King Faisal University in Saudi Arabia.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. None of this work has been published before submission

Date: 19/11/2020

Acknowledgements

I would like to thank and express my sincere appreciation to all those who have helped, assisted with and contributed to the successful completion of this thesis. Without their help and support, this journey would not have been possible.

First of all, I would like to thank my supervisor Dr. Gary Kinchin for his unconditional support and valuable guidance during the course of this study. I would also like to thank my second supervisor Professor Daniel Muijs, who offered wise comments.

I wish to express my gratitude to the lecturers, professors, staff, and the Dean of Education College at King Faisal University for their support and help during the data collection phase. The generous efforts of the participants in this study are greatly appreciated.

In addition, I would also like to thank the Ministry of Higher Education in Saudi for granting permission and facilitating the venues to conduct this study

The author is grateful to the Government of Saudi Arabia represented by King Faisal University for providing the necessary financial support, which allowed this research to continue unimpeded. The support is gratefully acknowledged.

My sincere gratitude and heartfelt thanks go to my mother for her understanding and encouragement.

Last, but not least, I am extremely thankful to my wife, my daughters, and my sons for their continued love and support.

Chapter 1   : Introduction

1.1   Background

1.1.1   Higher Education Context

In the modern era, Higher Education has witnessed a number of changes as a result of globalisation and its consequences. These repercussions have manifested in several aspects that have led to Higher Education emerging from its local isolation to open up to the countries and peoples of the world. The process of globalisation has thus forced universities to confront the new reality of universal participation and multiculturalism (Alameri, 2017).

According to Alameri (2017) and Stromquist (2007), internationalisation is considered a key contemporary trend in Higher Education, and it has become one of the most commonly used terms in universities as a major factor linked with the effects of globalisation. De Wit et al. (2015), in a review of Knight's (2003, 2004) well-known definition of internationalisation, describe it as: 'the intentional process of integrating an international, intercultural or global dimension into the purpose, functions and delivery of post-secondary education, in order to enhance the quality of education and research for all students and staff, and to make a meaningful contribution to society' (De Wit et al., 2015, p.29). As such, it is not surprising to read or hear that there exists so-called ranking of universities based on their level of internationalisation (Spencer-Oatey & Dauber, 2019).

Higher Education institutions face many challenges due to accelerated social and economic changes on both the local and the international level, the intensity of competition in terms of the appropriateness of university graduates for the labour market, and attention to the quality of university outputs (Alshamlan & Al-Fawzan, 2017). This has resulted in fierce competition between universities in order to attract the most outstanding students, while

there is an effort to review and reform the university systems, which benefit the graduates and their ability to adapt to and meet the requirements of the labour market (Ibid.).

Similarly, some European countries began reforming and developing their Higher Education systems some time ago. One of the most important reforms was the decision to directly link universities to the labour market, which has made them somewhat less dependent on government financial support (Kaneko, 2007). One of the most important contributing factors to the prosperity of these universities is the relevance and quality of their promotional products to meet the expectations of their consumers. For example, in recent years, Higher Education institutions in the United States have started to act as a Higher Education market, possibly due to the increased demand for high quality education (Alsalem, 2017). The British government's policy has also included paying attention to the expansion of Higher Education to increase participation and to encourage the creation of a more educated workforce (Molesworth et al., 2011). As noted above, this has resulted in fierce competition between Higher Education institutions due to an increase in the number of students as end users and Higher Education institutions' resulting desire to meet "consumer requirements" and capture the market (Ibid.). In other words, universities sell or offer products to consumers (students) and students decide on the type of product they prefer and whether or not these are suitable for the labour market.

Supporters of marketing Higher Education believe that this process will transform Higher Education into a more flexible and effective educational resource (Dill, 2003). In addition, they argue that it will provide better value for universities if they become more effective and responsive to the needs of society and the economy in general, and to the students and their parents in particular (Molesworth et al., 2011). According to Tomlinson (2018), "[t]he marketization of Higher Education is reported to actively encourage students to act as rational investors, informed choice-makers and indeed consumers, of their education" (p. 714).

In contrast, as stated by Kirdar (2002), it seems that university graduates in many Arab countries have not been as fortunate due to poor qualifications that prevent these job candidates from starting a career in their area of expertise and fitting into the labour market. It could be that the education system in many Arab schools and universities has always been more dependent on traditional learning and assessment approaches that focus on rote learning and memorisation than other more recently developed learning approaches. Such approaches might be developing critical thinking skills, problem solving, creativity, and using learning platforms, and one must also not neglect the failure of universities to meet labour market requirements (Al-Omro & Degri, 2017; Rugh, 2002).

In Saudi Arabia, there have been calls for major reform to Higher Education to match Saudi Arabia's position in both the Islamic and the international world. Some of the proponents calling for such reform of the Higher Education system in Saudi Arabia have included Al-Hazmi (2003), Elyas (2008) and Al-Miziny (2010). These researchers had a common agenda: the pressing need for urgent measures to reform the education system in general. However, the majority of these claims were based on curriculum reform rather than involving all aspects of education, including equal opportunities, and, of particular interest to this thesis, operating robust admission criteria, and using appropriate assessment methods that lead to strong outcomes for undergraduate students. The Saudi Ministry of Education has already embarked on several projects in this area, as in the case of the King Abdullah Education Development Project (TATWEER). Unfortunately, most of these initiatives have been designed to change the content of textbooks or focus on public education rather than including student-centred teaching methods such as cooperative learning, thinking critically and adopting methods of alternative assessment (Alnahdi, 2013).

More recently, Saudi Arabia has been looking to reform, largely in the social, political and educational spheres. Indeed, the new crown prince, Mohammed Bin Salman, is engaging with leaders around the world to this end. Saudi Arabia has also put forward its *Vision* and

ambition for the next 20 years: *Vision 2030*. One of the most important objectives of *Vision 2030* is the inclusion of five Saudi universities in the list of the top 200 universities in the world by the end of 2030. This objective may potentially create a competitive environment between Saudi universities to develop the educational system and the outputs of each university. Equally, and perhaps one of the most important reasons for this, is an expectation that this would lead to the desired improvements in university performance and the development of a coherent and aligned testing and assessment programme (Ali, 2017). One of the benefits of such a coherent and aligned testing and assessment programme is that it allows universities and policy makers to monitor standards and obtain evidence about its ability to predict the future performance of students (intended to monitor the universities). There should also be clear admission criteria characterised by equal opportunities and integrity. The performance of Saudi universities should meet the objectives of *Vision 2030* in building Saudi society, including the development and assessment of higher-level skills among learners, as well as the appropriateness of the output for the labour market. A more detailed analysis of *Vision 2030* appears in the next chapter.

1.1.2   University Admission and Student Performance

Universities face numerous challenges each year, including the process of making decisions concerning the admission or otherwise of applicants (Tesfa, 2013). This process can be compounded by competition, an increase in student numbers and a decrease in university resources predicting future academic success (Alghamdi, 2007). Educators and admission officers together seek to decide what contributes to the success of learners regardless of their discipline or area of study within universities. Administrators who are responsible for the admission policy at a particular institution need to be accurate and objective when making

such decisions using suitable admission standards, which then support them in reaching a decision characterised by equity, accuracy, and objectivity (Alsaif, 2005).

Researchers in the fields of psychological and educational measurements in scientifically developed countries continue to address this problem. Consequently, much research has been conducted employing a variety of objective tests in order to assist decision makers in predicting the future achievement levels of students applying for university education. A broad range of criteria are used to inform decisions to offer learners places at universities (Al-Hattami, 2012; Alnajjar, 2001; Brown et al., 2002) Researchers have explored these criteria and offer predictions for the learners' success at the academic level (Hill, Storch, & Lynch, 1999; Rumsey, 2013). Determining accurate criteria that are tightly linked to learners' success remains a complex task, and research has led to a variety of outcomes and findings. Interest has centred on cognitive factors (e.g., SAT) (Camara, 2008; Rothstein, 2005), A-level grades (Bourner & Hamed, 1987), or non-cognitive factors (e.g., race, ethnicity, location) (Wang, 2003). A review of this research indicates that High School grades and standardised tests scores, such as Aptitude Tests or Achievement Tests (Furnham & Monsen, 2009; Gill, 2007; Messick, 1994), and demographic characteristics (Camara & Echternacht, 2000; Geiser & Studley, 2001; Shaw, et al., 2012) are significant predictors of student success at university level (Al-Dossary, 2001; Aldoghan, 1985; Bourner & Hamed, 1987; Camara & Echternacht, 2000; Geiser & Studley, 2001; Kobrin et al., 2008). Many of these studies have addressed the most common standardised tests such as the SAT (Scholastic Aptitude Test), and A-level tests, and have explored the relationship between test scores and academic success at university. In general, the SAT score was found to be a significant predictor for learners' academic success in Higher Education (Al-Dossary, 2001; Camara & Echternacht, 2000; Geiser & Studley, 2001; Kobrin et al., 2008). Although the A-level test is widely recognised in UK universities, many researchers have presented different outcomes concerning the link between A-level scores and academic success. To

illustrate, some studies have found either no significant link or a negative relationship between A-level scores and academic success in Higher Education (see Peers & Johnston, 1994). Other studies have reported small but positive correlations (Bourner & Hamed, 1987). SAT and High School scores were found to correlate very positively with academic success in the first year of university (Kobrin et al., 2008).

Some studies have focused on the relationship between performance in specific subjects at High School and success in pre-university programmes. Many studies have identified significant correlations between certain subjects at High School, such as English, and success in the pre-university exam that is used as a measure for entrance to university (Alavi, 2012). Research has also focussed on the International English Language Testing System (IELTS) as a predictor of performance and success in some subjects at university (Dooey & Oliver, 2002). For example, Dooey and Oliver (2002) found that there was a positive correlation between IELTS reading scores and the success of business students, although this was not found to be the case with science and engineering students. In contrast, other researchers reported no positive correlation between IELTS scores and the academic outcomes of International students (Cotton and Conrow, 1998), while Hill and colleagues (1999) attempted to compare scores on IELTS and the Test of English as a Foreign Language (TOEFL) as predictors of academic success. A strong relationship between IELTS and GPA at college and a weak relationship between TOEFL and academic success were reported. Further studies have focussed on additional relationships and have addressed the predictions of these test scores in certain majors (Murshid, 2013; Wang, 2013; Wilkinson, Zhang, & Parker, 2011) for which admission standards are considered competitive. Such studies are very useful in facilitating reform to policies that focus upon admission and the setting of appropriate educational standards for learners.

The process of making decisions concerning the admission or otherwise of applicants is also a matter for Higher Education within the geographical region in which this thesis is based

(the Arabic Gulf). Some studies in the wider international literature are comparable with respect to cultural and educational characteristics within Saudi Arabia. Here researchers have focused on studying the predictive value of the percentage of High School and university academic success (Albanai, et al. 2004; Almiqdadi, 2010; Almokhlafi, 2001; Darwaza, 1987). These studies found that High School assessment scores are the best predictors of future academic success. However, there are some researchers in Arab countries who have reported that High School achievement was not a sufficient predictor for success at university or that it only exhibited a weak relationship to Grade Point Average (GPA) at the university-level (Alkhattab, 1989; Rahal, 1987). These studies confirmed the connection between high school overall percentages and university performance, especially in early years and based on the accumulative average.

In Saudi Arabia there is limited research which has explored the correlations between standard ability or High School Achievement Tests and later success at university (see Alqatei & Alharbi, 2012; Alshumrani, 2007). One possible explanation for this absence could be that, for some time, Saudi universities have only depended on High School scores for university admissions, and, on occasion, on personal interview and/or local placement tests and, even then, only for certain majors. As an example, one university (King Fahd University for Petroleum and Minerals) has operated a standardised test for ability known as the RAM1 (Aptitude Test) and the RAM2 (Achievement Test) (Aldoghan, 1985; Alsaif, 2005; Attia, 2002). These researchers have found a significant correlation between RAM1 and 2 as well as High School performance and success in the first year of university education. However, few studies have investigated this test as a predictor for overall university success.

---

[1]More information about RAM1 and RAM2 can be found in Chapter three.

1.1.3   The National Centre for Evaluation and Measurement (QIYAS)

In 2001, The National Centre for Evaluation and Measurement was established in Saudi and, therefore, standardised tests were applied in the national universities. This may be another reason for the identified lack of literature linking Aptitude Tests with academic success in the Saudi context. A close examination of the literature revealed that five studies have been undertaken to address the predictions of the standardised tests published by the Saudi Arabia National Centre for future academic performance (Alshumrani, 2007; Alshehri, 2011; Alqatei & Alharbi, 2012; Murshid, 2013; Alkushi & Althewini, 2020). Each study was found to have its limitations (see section 3.5.9), and this work, in turn, assisted in developing the research undertaken for this thesis. Generally, these studies were either restricted to certain populations, to certain majors, or both (this is discussed in the following section). Thus, an obvious gap exists, and studies in this area would significantly contribute to knowledge, aiding policy makers in the field of university admissions in developing and offering clear and transparent information regarding admission standards. In spite of the importance of admission standards in making decisions about student admission or rejection, the unavailability of sufficient evidence for their validity, reliability, and objectivity makes unquestioning dependence on their results somewhat misguided. This therefore encourages studies into the predictive value of these standards for future academic success to determine the degree of admission standards' validity and adequacy for making decisions regarding student admission or rejection.

1.2   Saudi Arabia: Admission to Higher Education

In Saudi Arabia, Higher Education Institutions use High School percentages as the main criterion for admission. Often, the minimum score is 80%. Most colleges and universities

have traditionally taken this criterion as the only admission standard for accepting students into Higher Education. However, one institution has introduced a content-specific exam or interview as a second criterion used to accept or reject students (Almarshad, 1996). King Fahd University for Petroleum and Minerals was the first university to employ an aptitude test and Achievement Test, as mentioned above, in conjunction with a high school degree in 1997 (Musa, 2003, p.3).

There exists a paucity of studies addressing the issue of the predictive validity of the admission standards applied in Saudi universities. This is especially true considering that QIYAS was only launched in 2010. Research in this area has not been forthcoming. Indeed, following a review of the relevant literature, only three studies were found to examine the issue of academic success in relation to High School scores, Achievement Tests and Aptitude Tests. Those studies, as mentioned in the previous section, yielded different outcomes and limitations regarding the targeted gender population, the majors covered in the investigation, and the research design and context. The range of outcomes and limitations point to a clear need for further investigation regarding which of the three admission criteria most strongly contribute to students' academic success at undergraduate level. Given the state of this area of research, additional features have been included in the present thesis. Firstly, no study has been found to address predictive validity for students who changed their major after the first year. Secondly, none of the studies reviewed had covered the discipline of education in any of the Saudi universities. Thirdly, no research in Saudi Arabia has addressed the relative performance of males and females in Higher Education. Fourth, no studies have employed both qualitative and quantitative data collection. Therefore, investigating the predictive validity of academic success that includes both males and females would be of great interest to researchers and policy makers within the Arabic region, particularly given the social and cultural changes taking place at a national level, and this against the backdrop of *Vision 2030.*

Being an instructor and a member of staff at the Educational College at King Faisal University (KFU), the author of the present study became more interested in the admission standards at the Educational College, since many students tend to change their major during their university study. Further, from experience of teaching courses within the Education degree, it was observed that some students with high scores in High School and QIYAS tests often fail some of the modules within the programme. This gives further importance to investigating what contributes significantly to students' success at the Educational College at KFU and what, if any, impact this might have on the ways in which undergraduate education students are assessed. This strengthens the need for investigating the predictive validity of the admission standards in a School of Education to explore which of the criteria for admissions is the best predictor for academic success, and, also, for seeking the views of lecturers on these predictions.

1.3   The Purpose of the Research

The purpose of this study is to investigate the predictive validity of the current admission standards applied at the College of Education at KFU in Saudi Arabia and explore which score among the current criteria is associated with the strongest contribution to students' future academic success. Furthermore, this study attempts to include students from both gender groups, and very few studies have included both genders at the general level. To the best of the author's knowledge, no such studies have been completed in the Saudi context. This study aims to explore any possible variation between the criteria items in terms of gender groups in Saudi. Moreover, since none of the previous studies have addressed the issue of students changing their major after initially being admitted to certain majors at the university, this study is intended to explore the academic performance of students who change their major after already starting their university education.

It should be noted that it is in the interests of the administrators in the admission department at KFU in general, and the College of Education in particular, to investigate students' performance with respect to their abilities. As noted above, many students who scored highly in QIYAS tests, as well as in their secondary tests, struggle in some Education modules and units on the BA programme, but it is not yet clear why this should be. This study, therefore, includes a qualitative dimension by interviewing a number of key academics within the KFU Education department to seek their opinions regarding the relationship between the assessment undertaken at the end of High School and that within the College of Education at KFU. Lastly, since the launch of the National Centre of Measurement and Assessment and the application of QIYAS tests in 2002, the centre has, over time, become increasingly interested in the relationship between QIYAS scores and student performance across all majors in all Saudi universities. Therefore, this study is intended to provide useful analysis regarding students' academic performance in relation to QIYAS tests in the case of one institution.

1.4   Research Questions

The current study addresses the following main research questions:

1.   Is there a statistically significant relationship between the students' accumulative assessment scores at the end of High School and the accumulative rate in the College of Education at the end of 2014 at King Faisal University in Alahsa, Saudi Arabia?

2.   Is there a statistically significant relationship between the student mark in the General Aptitude Test at the end of High School and the accumulative rate at the College of Education at the end of 2014 at King Faisal University?

3.   Is there a statistically significant relationship between the student mark in the Achievement Test at the end of High School and the accumulative rate at the College of Education at King Faisal University?

4.   What is the degree of the following independent variables' ability in terms of interpreting the accumulative rate for the department within the College of Education at King Faisal University?

-      The accumulative rate in High School.

-      The marks in General Aptitude Tests.

-      The marks on the Achievement Test.

5.   Which admission standards have the greatest ability to predict success at the College of Education at King Faisal University?

6.   Is there a statistically significant relationship between the score on the General Aptitude Test, Achievement Test, and High School percentage with respect to the variables of gender and change of major?

7.   What are the College of Education Lecturers' perceptions about the current admission standards used at the College of Education at King Faisal University (The accumulative rate in High School, the marks on the General Aptitude Tests, and the marks on the Achievement Test).

8.   What opinions do the College of Education Lecturers have about the relationship between assessment undertaken at the end of High School and that undertaken within the College of Education?

1.5   Research Hypotheses

With respect to the above research questions, this study proposes the following hypotheses:

1) There is no statistically significant relationship between the students' accumulative rate in High School (SGPA) and the accumulative rate in the College of Education (UGPA) at King Faisal University.

2) There is no statistically significant relationship between the students' scores in the General Aptitude Test (GAT) and the accumulative rate in the College of Education in King Faisal University (UGPA).

3) There is no statistically significant relationship between the students' scores in the Achievement Test (ACT) and the accumulative rate in the College of Education in King Faisal University (UGPA).

4) There is no statistical variation between the following independent variables in interpreting the UGPA in each department within the College of Education at King Faisal University:

   (a) The accumulative rate in High School.

   (b) The marks of the General Aptitude Tests.

   (c) The marks of the Achievement Test.

5) There is no statistical variation between the admission standards in predicting student success at the College of Education at King Faisal University.

6) There is no statistically significant relationship between the score of the General Aptitude Test, Achievement Test, and High School percentages considering the following variables: gender and change of major.

1.6   Contribution to Knowledge

The introductory chapter has identified several gaps in the research within the area of the predictive validity of the admission standards applied in Saudi universities and which predictor is most strongly linked to students' academic success. This study is expected to offer useful and additional insights with a focus on one particular Saudi institution.

The thesis, in part, is also intended to contribute to current debates around particular aspects of *Vision* 2030 in terms of evaluating the criteria used in the College of Education at KFU and examining their validity and ability to predict students' academic success. Identifying the strengths and weaknesses of the admission process and the criteria upon which educational institutions were based and the extent to which these tools have been used to accept students in predicting academic success can be of significant importance for studies related to academic achievement prediction or academic success (Morales-Vives et al., 2020; Makransky et al., 2017). Recent studies point to an awareness among teachers and officials in admission and registration about whether these standards are related to students 'performance in college (Alkushi & Althewini, 2020).

It is also worth mentioning here that predicting academic achievement contributes to providing education for learners when assessment is not feasible. Given what is happening at the moment during the coronavirus pandemic (COVID-19), at some point this year (2020), students in some and high schools, as well as universities, have not been able to take their end-of-school exams. Therefore, most teachers have used past assessments as a predictor of future performance for students and are required to indicate what their likely performance may have been in these summative assessments.

It can also be said that the importance of predicting students 'future performance is related to results, gains, and risk reduction. for instance, student success plays a vital role in educational institutions, as it is often used as a metric for the institution's performance. Early

detection of students at risk, along with preventive measures, can drastically improve their success (Alyahyan and Düştegör, 2020). This is what KFU seeks to achieve by applying admission criteria for students who intend to enroll in its colleges. These predictive studies allow the university to produce strong evidence on its ability to understand whether students who have passed the admission criteria used in the university will thus reach success in the various university programs, as well as determining the extent to which their standards are predictable and which of them are required for review and assessment.

This research may contribute to bridging the gap between the Saudi labour market needs and higher education outcomes. This is done by clarifying the ability of the admission criteria used in the College of Education at King Faisal University to predict academic success. It is also achieved by explaining the methods that are used in evaluating learners within the College of Education, the skills and capabilities achieved by those methods, and the extent to which all of these are related to the needs of the labour market and the *Vision 2030* project, which seeks to achieve an education that contributes to the progress of economic development, support of the national economy and investment in higher education.

In light of the global trend in the promotion of higher education, which the Kingdom seeks to join, the marketing of higher education considers university services and procedures of great importance to all stakeholders involved directly and indirectly with the university, whether students, employers, or the community at large, for whom the university seeks to meet a complex set of needs. While students are interested in investing well in terms of time and effort and acquiring knowledge, skills and specialisations that are compatible with the requirements of the labour market, employers expect to attract and retain college graduates who are able to master the work and solve problems encountered on the job. In the meantime, the community aspires for graduates that are able to cover different sectors. Nicolescu (2009) pointed out that the quality of university services is of utmost importance as it leads to gaining customer or stakeholder satisfaction and contributes to the marketing of university

education. Thus, this research tries to contribute to improving the quality of university services related to admission criteria, assessment of methods for learners, which contributes to the promotion and marketing of King Faisal University in general and the College of Education in particular in light of the fierce competition between universities.

This thesis also seeks to gather the views and attitudes of faculty teachers at the King Faisal University College of Education towards current assessment methods used in both public schools and in Higher Education, and whether these are appropriate to start alongside other the developments in the field. Such assessment methods are also scrutinized to ascertain whether they can measure the skills and knowledge that are consistent with the modern age and globalization, as well as the Saudi Arabian labour market. *Vision* 2030 addresses the historic absence of any comparison between the relative performance of boys and girls in universities in Saudi Arabia, which, in turn, has resulted in no studies comparing performance on the basis of prediction. This thesis gives attention to such a deficit.

The outcomes from the proposed research will provide useful policy and practical implications for neighboring colleagues in both similar and different academic majors, Colleges/Faculties within other universities in Saudi Arabia specifically, and within the Gulf region as a whole, and further beyond more generally.

1.7   Structure of the Thesis

This research consists of eight chapters in addition to the introduction. The second chapter, which refers to the research context, includes a discussion of the educational context in Saudi Arabia, the assessment system used in public education, and an overview of the Centre of Measurement and Assessment in Education (QIYAS). It concludes by setting out the criteria used to accept students at King Faisal University. The third chapter contains a literature review and aims to offer a comprehensive understanding of the subject of the current

research, including predictions of academic success in pre-university tests and the literature that has touched upon this topic. The fourth chapter of this thesis contains a research methodology and explains how this research has been conducted in terms of sample selection, methods of data collection, and analysis. The fifth chapter is dedicated to the experimental sample (pilot study) that was applied to the interview, which included doctoral students at Southampton University who specialise in linguistics and have a background in Saudi education and the methods of assessment used in it. Chapter 6 contains the results of the current research, the quantitative method and its response to questions 1 to 6 using SPSS software program Version 25, and how the open interview was used with the lecturers from the Faculty of Education at King Faisal University to answer questions 7 and 8. In Chapter 7, the results are discussed with respect to the literature. The final chapter contains the conclusion, including the limitations of the results and the contribution to the current research in terms of knowledge, the expected impacts of this research, as well as the recommendations for future research based on the current findings.

Chapter 2   The Research Context

2.1   Introduction

Chapter One set out the background to the subject of the predictive value of some standards employed for admission to universities in different countries, and some Saudi universities in particular. Additionally, it briefly outlined the aim of the research and provided the research questions and assumptions relevant to this study. The introduction chapter was concluded by addressing the contribution of this research to scientific knowledge. This chapter addresses the public education system in the Kingdom of Saudi Arabia and the authorities supervising it. Moreover, it highlights the relevant types of High School, school subjects and assessments. This chapter concludes with a brief description of the admissions system at King Faisal University at Alahsa, since this study concerns bachelor students from the College of Education at this institution.

2.2   The Saudi Education System

2.2.1   The Saudi Education System: A brief history

In Saudi, the government is primarily responsible for education, providing it free of charge for all students and at all stages. In addition, the government pays a monthly remuneration for students from some schools, such as Qur'an memorisation schools, where monthly payments for elementary school students are SR 250, approximately equal to £42, SR 500 for intermediate students, approximately equal to £84, and SR 600 for high school students - approximately £100 per month (Aljuraid, 2013). According to the financial affairs regulations regarding universities (item 41), a monthly remuneration is paid to students of colleges and universities for undergraduate study at governmental universities of SR 850,

which is equivalent to £141 per month for specialties at theoretical college and SR 1000 for students of scientific colleges, which is almost equivalent to £168 (King Saud University, 2019). The government is also responsible for providing all equipment, buildings, textbooks and teaching aids, in addition to the training of teachers and their employment, securing their salaries, and scaffolding their promotions to top jobs in education. The government also bears responsibility for pensions (Ministry of Education, 2004). Saudi education is supervised by a range of sectors which are as follows:

2.2.1.1 The Ministry of Education

This is the first and oldest supervising body within the Saudi education system (Secondary-General-Technical). Since its establishment in 1926 in the name of the Directorate of Knowledge, it has supervised education at all levels for both boys and girls. The body was initially called the Directorate of Knowledge, but this was then amended to the Ministry of Knowledge in 1952, and then amended again to the Ministry of Education in 2002 and 2015. After the establishment of the Ministry of Higher Education, the General Presidency for Girls' Education, Technical and Vocational Training Corporation became the Ministry of Education but only supervising general education (primary, intermediate and secondary) (Alhammed et al., 2007).

2.2.1.2 The General Presidency for Girls' Education

This body has had oversight of public education for girls since its establishment in 1960 when it supervised the Colleges of Education for girls. In 2002, the General Presidency for Girls' Education was incorporated into the Ministry of Knowledge. The name was then changed to the Ministry of Education (Alaqeel, 2013).

2.2.1.3   The Ministry of Higher Education

This body has supervised Higher Education since its establishment in 1975, although Higher Education started before its establishment, e.g., the King Saud University was established in 1958 as the first Saudi university (Alsunbol et al., 2009). The Ministry of Higher Education was incorporated into the Ministry of Education in the year 2015, and then the Ministry of Education became the supervisor of public education and Higher Education, with the exception of technical education and vocational training.

2.2.1.4   The Technical and Vocational Training Corporation

The Ministry of Education supervised Technical Education and the Ministry of Labour, and Social Affairs supervised vocational training. In 1980, the Technical and Vocational Training Corporation was established. This corporation became responsible for the implementation of plans and programmes for the preparation and training of technical and professional personnel. It also supervises the industrial institutes and secondary technical education institutes (Alhogail, 2011). Saudi public education thus consists of three stages that are the Primary Stage, with a study period of six years of age (6-12) years and admitting those children at 5 years old and 6 months dependent on completion of kindergarten. Intermediate: there is a study period of almost three years (13-15) years and this starts from the awarding of the sixth primary class certificate. Secondary: there is a study period of almost three years (16-18) and this begins after the awarding of the third intermediate class certificate (Ministry of Education, 2010).Education in the Kingdom of Saudi Arabia is not considered mandatory. This means that even if a child is not enrolled in any stage of education, the guardian is never punished. This is in spite of the fact that the Saudi education policy, in Article 10, states the following, "the pursuit of education is obligatory on every individual by virtue of Islam, and the dissemination and facilitation of education in various

stages is the duty of the government as much as its ability and potentiality" (Hakim, 2012). The Saudi education system separates girls and boys at all educational levels, except in nurseries and kindergartens (Education Policy, 1995).

2.2.2   High School

2.2.2.1   Overview

High School education in Saudi Arabia covers three years. It follows on from middle school. The secondary stage contributes, similarly to the other educational stages in Saudi Arabia, to the achievement of the general objectives of Saudi Arabia, in addition to the achievement of the following specific objectives:

1-Achievement of loyalty to Allah Almighty and the undertaking of actions pure to Allah.

2-Achieving loyalty to the general Islamic homeland and the private homeland in particular (KSA).

3-A pledge to students' abilities and different aptitudes appearing during this stage and their direction in accordance with achieving the objectives of Islamic education in general.

4-The development of scientific thinking of the student and enrichment through research and an experimental spirit.

5-The establishment of positive awareness when facing different thoughts and ideas (Alhogail, 2011).

The Saudi Scientific Institute is considered one of the first schools to be concerned with education beyond the preliminary stage. This institute, in which holders of the preliminary certificate or its equivalent were accepted, was opened in 1927. Study at the institute lasted

only three years, preceded by one preparatory year. In 1947, the study period at the institute increased to five years and at the end of this time, the student obtains their secondary certificate.

Since its establishment in the Kingdom, High School was the focus of a series of laws and developments including changes in the study system and amendments to the curricula. Development and diversification were undertaken in secondary majors. Accordingly, commercial, agricultural, technical and other types of High Schools were established (Alghamdi and Abduljawad 2005).

2.2.2.1.1 Types of High School systems

Over time, a number of different schools have existed in Saudi and these were:

2.2.2.1.1.1 Traditional secondary

This system was adopted by the Kingdom's schools since their establishment in 1936. It was a system similar to that in other Arab countries in which traditional subjects were taught (complementary to the intermediate stage) such as the Arabic language, religion, mathematics and sciences. The student is taught all the subjects in the first class of secondary. For the second and third secondary classes, the student selects either the Arts division or the science division. A graduate of this division holds the High School certificate. This certificate often qualifies the student to enter governmental or private universities. Traditional education in the secondary stage, for more than fifty years, remained assigned to the Kingdom's High Schools for both boys and girls (Alsunbol et al., 2009).

2.2.2.1.2 The comprehensive High School

The application of this system came in response to the recommendation of the state's second national development plan (1975-1980), which determined the start of comprehensive education at the secondary stage. The educational system of the comprehensive school

featured a number of advantages that made it different from the traditional secondary (Alhammed, 2005), including:

The student selects the subjects they wish to study, and sets their own study schedule, as well as the number of hours they wish to study. This system is similar to the university system. The assessment often occurs through use of the short tests or continuous assessment. Summer study exists, where the student may opt to study during the summer and, therefore, may shorten the study period from three years to two-and-a-half. The total study period in this comprehensive secondary is more or less six-study classes according to the abilities of the student. The following was required for the completion of comprehensive secondary study: the student must have had 120 credit hours in the study subjects chosen by the student within the limits of the departments and sections, and 30 hours credited hours in one or more of the activities provided by the school.

The school includes sections from which one will be chosen by the student upon joining the school. The sections include: the religion section, the languages and social sciences section, sciences and mathematics, technical sections and the general studies section (Educational Documentation - Ministry of Education, 2004). In 1991, the resolution of the Council of Ministers (number 105) was issued for the cancellation of this system with a return to the previous system with compliance to branching High School into four departments starting from the second secondary class and, in addition, the restudying of the curricula and their organisation was decided accordingly.

Thus, High School had returned with some amendments to the old system. The Ministry of Education made amendments to some sections and specialties such that there were four sections starting from second class from which the student selects one. They were legal sciences (Arts previously), natural sciences (the scientific section previously), the administrative and social sciences section, and the technical sciences section. In addition,

new curricula were added which were taken from the developed system, the most remarkable of which were the curricula relating to computer sciences, the library and the research curriculum to be added to the curricula existing in the traditional system. The traditional system continued to be adopted in Saudi until very recently.

2.2.2.2 Types of High School

There are three main types of High School in Saudi Arabia, which are: Holy Qur'an Memorisation, Scientific Institutes, and General High Schools.

2.2.2.2.1 Holy Qur'an memorisation High School

These types of High Schools are accountable to the Ministry of Education. The first school of this type was open in Makah between 1967 and 1977. The study period in this school comprises three years, focusing on Qur'an studies, Islamic sciences, and the Arabic language. The rate of its total class periods over the three years reaches 78% of the total class periods. The student is also taught history, geography, English language and physical education. The students of these schools are given a bursary of 600 SAR per month, roughly equivalent to £100 (Higher Committee for Education Policy, 1982).

2.2.2.2.2 Scientific Institute High Schools

Imam Mohammed Bin Saud Islamic University supervises this type of High School and these institutes focus on the legal (Sharia) sciences and Arabic language sciences, with 80% of the total number of class periods dedicated to these and 20% for the social and English language subjects. The first scientific secondary institute was opened in 1951 in Riyadh (Curricula of Secondary Stage in Scientific Institutes, 1991).

2.2.2.2.3   General High Schools

This type of school is very common in the Kingdom of Saudi Arabia and represents 94% of the total High Schools for boys and 95% of total girls High Schools. The inputs of this type of High School experienced quantitative development in all aspects (number of students, teachers, number of schools, etc.). Table 2-1 highlights this development: (Education Statistics in the Kingdom of Saudi Arabia - Ministry of Education, 2004).

Table 2-1: The development in High School for Boys and Girls

Years	Schools (NO)		Classes (NO)		Students (NO)		Teachers (NO)		Administrative (NO)	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
95/1996	907	918	7755	7375	203742	202016	14598	16970	811	2037
14/2015	2128	2199	13179	15411	412906	417935	36373	50308	2979	7458

From the above table, it can be seen that there has been an increase in the number of High Schools for both genders with growth of (134.6% for boys' schools and 139.5% for girls' schools) between the school years 1995/1996 and 2014/2015. The same applies to the study classes for both genders. This increased by 69.9% for boys' classes and 108.9% for girls' classes, while the rate of the number of male students and female students reached 102.6% and 106.8% respectively. Moreover, the rate of male and female teachers increased by 149% for males and 196% for females. The administrative staff and departments increased by 267% for males and 266% for females. Therefore, from the previous statistics, we note an acceptable quantitative development.

In the first school year, both male and female students are taught general curricula and the specialty starts during the second and third class, arts and scientific for girls and legal and Arabic sciences, administrative and social sciences, and natural sciences for boys.

2.2.3   The Assessment system in Public school

The student evaluation process in public Saudi education has undergone a number of changes and modifications which can be summarised as follows:

2.2.3.1   Procedure for the years 1926-1941

Evaluation for students was through written examinations at the end of every three-month period for all Saudi schools and in all subjects. Qur'an, poems and reading are excluded from these subjects, where they are only evaluated orally. A mark of 50% is required for success in any subject to obtain the minimum pass rate (Ministry of Education, 1998).

2.2.3.2   Procedure for the years 1954-1960

In 1959, the first system of examination issued included the marks for subjects, the organisation of the mechanisms of examinations, and their unified dates in the Kingdom of Saudi Arabia. In the first and second year of elementary school, all subjects were evaluated orally. From the third year of elementary to the third year of secondary, learners are evaluated in writing except for the Qur'an and reading. The learner is required to obtain 50% of the grand total for each item of success; examinations are at the end of each semester (there are two semesters in each academic year) (Al-Ghamdi et al., 2005).

2.2.3.3   Procedure for the years 1961-1997

In this era, evaluation examinations for learners in the third secondary year were designed and prepared by the Ministry of Education. This was known as so-called question centralism. To clarify, this means that the questions are designed by the Ministry of Education for the second semester for only students in the third secondary year. The questions are then sent to the school principals in sufficient time. It is forbidden to open the examination questions before the day of each subject examination. Then, after learners answer the questions, they are sealed in envelopes and returned to the ministry. A group of committees formed by the ministry corrects the answers. Finally, the results of students are published in the official newspapers after being reviewed. However, after the emergence of Internet technology, the results have been announced via the website of the Ministry of Education. Marks for each subject are distributed as follows: 30% of the total mark is from work undertaken during the semester (short examinations, participation and class activities). 70% of the total marks come from the final examination at the end of each semester (Ministry of Education, 1998). At the end of 2007, question centralism was halted. Every subject teacher became responsible for the design of questions for their students (www.moe.gov.sa).

2.2.3.4   Procedure of continuous evaluation for the elementary stage in 1998

In this year, a new set of regulations regarding the evaluation for the first, second and third elementary year students came into force. Continuous evaluation applied, in which the success of any student depends on the mastery of basic skills in the subject. The learner is given a number from 1-4* to reflect his academic level; the evaluation of students takes place during the school year through short examinations, participation, and homework. There is no examination at the end of the semester as there was in the past. In 2006, this type of evaluation was applied to all elementary school students and is still in place today.

* (1) means that the student has mastered all skills in the prescribed subject.

(2) If a student has mastered 66% or more of the evaluated skills.

(3) If the student has mastered at least all the skills to the minimum threshold.

(4) If the student has not mastered all the skills to the minimum threshold. (This means that the student is promoted to the next year but is referred to the Committee of Direction and Guidance to make a final decision).

2.2.3.5   New evaluation regulation 13 for Intermediate and High School

The most significant features of this new evaluation are as follows:

1. Allocate 100 marks for each subject per semester (academic year is divided into two semesters).

2. The academic achievement of students is evaluated by:

A: Semester works: each subject is worth 50 marks and this is divided into 30 marks (attendance, participation, homework, research or reports). There are 20 marks for the two examinations prepared by the teacher during the semester. The Qur'an is examined orally from 100 marks, from which 5 marks are awarded for the regularity of the student's attendance.

B: 50 marks for the examination of the end of the semester.

3. The learner is considered to have passed the subject if he obtains the minimum threshold (50%) and provided he obtains 20% of the final examination each semester.

4. The final examinations for subjects unfold according to the following sequence:

A. First session examination: is the examination at the end of the academic level and is for all students.

B. Second session examination: an examination of the subjects not passed by students in examination of the first session.

C. Failed subject examination: an examination of the subjects not passed by students in examinations from the second session. This demonstrates that the learner has the opportunity twice, in the event of failure, to pass each academic subject.

5. The Special Grade of Subjects is granted as in the following table:

Table 2-2: The meanings of Intermediate school certificate symbols

Mark	Grade	Symbol
95-100	Upper Excellent	A +
90-94	Excellent	A
85-89	Upper Very Good	B +
80-84	Very Good	B
75-79	Upper Good	C +
70-74	Good	C
61-69	Upper Pass	D +
50-60	Pass	D
less than 50	Complementary (Not passed First Session)	E +
	Failed (Not passed Second Session)	E

6. GPAs calculated for Secondary Certificate are as follows:

A. 25% is calculated from the first and second level averages (the average of the first secondary year percentage).

B. 35% is calculated from the third and fourth level averages (the average second secondary year percentage).

C. 40% is calculated from the fifth and sixth level averages (the average third secondary year percentage). These percentages together constitute the Secondary Certificate average for the learner (Ministry of Education, 2014).

2.2.4   The National Centre of Assessment in Higher Education (QIYAS)

The only unified criterion for admission to universities in Saudi Arabia used to be the High School exams despite the fact that some universities would require a personal interview with the applicant, except for King Fahd University of Petroleum and Minerals, which used the Achievement and General Aptitude Tests known as RAM 1 and RAM 2. However, after the establishment of The National Centre for Assessment in Higher Education in 2002, a new standard was introduced and became compulsory for all universities and colleges in Saudi Arabia, and this includes the General Aptitude Tests, as well as the Achievement Test prepared by the National Centre for Assessment in Higher Education (QIYAS).

The QIYAS Centre was created solely to focus on admission to Higher Education and for the purpose of rallying efforts and the expertise of the universities to create a scientific basis for these tests. The National Centre for Assessment in Higher Education in Saudi Arabia was established to fulfil the following criteria:

1-   One of the requirements for admission into universities and colleges in Saudi Arabia is tests whose results are considered a criterion together with the High School examination. This test is carried out according to the following:

a)   A test to measure students' skills and attitudes.

b)   A test to measure academic achievement.

Such tests are standardised for disciplines that fall under one category (be it scientific and/or theoretical).

2-   Admission tests should be allowed to be offered many times throughout the year but be restricted to three per student.

3-   A financially and administratively independent centre should be established under the name of the 'National Centre for Assessment in Higher Education at the Ministry of Education'.

4-   A financial return should be acquired to cover the costs of holding these tests, including expenses related to the operation of the centre and its development, as well as the relevant research expenses (fees are 100 SR, which is equivalent to £22).

2.2.4.1   The Mission and Objectives of the Centre

The mission of the National Centre for Assessment is to achieve justice and equality of opportunity in Higher Education based on sound scientific grounds. Additionally, The Saudi National Centre for Assessment (QIYAS) aims to account for the following objectives:

1-   To play a leading role in the development of educational assessment tools in Higher Education.

2-   To contribute to raising the level of performance and efficiency in Higher Education.

3-   To prepare entry exams for Higher Education institutions.

4-   To provide advisory services to the assessment centres in the various educational institutions.

5-   To monitor scientific research and studies, and specialised research in the field of educational assessment.

2.2.4.2   Educational Testing

Educational testing is the first building block to be developed by the National Centre for Assessment with the assistance of a group of experts in this area, whereby the centre offers

some of the most important educational tests, including tests of general abilities and academic Achievement Tests.

2.2.4.2.1   General Aptitude Tests

This test measures the analytical and evidentiary ability of the student; in other words, it focuses on the ability of the student to learn regardless of their own proficiency in a particular subject. This can be done by assessing the following:

1.   The ability to understand (comprehension).

2.   The ability to perceive logical relations.

3.   The ability to resolve issues based on basic mathematical concepts.

4.   The ability to make inferences.

5.   The ability to make analogies (QIYAS).

The test is divided into two parts: a verbal part and a quantitative part. With regards to the verbal part, this consists of the following types of questions:

-    Absorbing reading materials: understanding and analysing the comprehension texts by answering questions on the content of these texts.

-    Completing sentences: understanding the wording of short texts and deducing what is needed for them to become useful sentences.

-    Verbal symmetry: recognising the relationship between a pair of words at the beginning of the question, and deciding which ones are similar from a list of multiple choices.

- Contextual error: understanding the overall meaning of a sentence and then selecting the word that does not agree in terms of meaning.

- Association and difference: recognising the relationship between three options and determining the irrelevant choice.

As for the quantitative section, this includes the right kinds of mathematic questions to test abilities according to subject or discipline in High School (scientific or theoretical). It focuses on analogy, deduction and problem solving.

The distribution of questions in the scientific disciplines is roughly distributed according to the following:

1/ Arithmetic questions: 40%. 2/ Engineering questions: 24%.

3/ Algebraic questions: 23%. 4/ Analytical and statistical questions: 13%.

As for the theoretical disciplines test, this contains questions that cover arithmetic, geometry and analysis.

2.2.4.2.1.1 Examples of test questions

1: The verbal section

Two words have a certain association, followed by a list of four options, with only one of these choices having a similar link to the two words stated at the beginning of the question. Choose the correct answer from the choices given below:

Watch: Time

A: Sun: Moon   B: Balance: Weight

C: Thermometer: Cold   D: September: October.

2: The quantitative section

An athlete is competing to perform 10 jumps. He would gain five points for every successful jump, and only two points for a failed jump. At the end of the competition, he collected 41 points. How many failed jumps were there?

A.   3   B.   5   C.   7   D.   10

2.2.4.2.2   The Academic Achievement Test

This includes High School graduates with two disciplinary subjects and those wishing to enrol into Higher Education faculties:

2.2.4.2.2.1   The Academic Achievement Test for scientific disciplines

The academic Achievement Test questions focus on general concepts in the scientific field, including biology, chemistry, physics and mathematics. Questions in the Academic Achievement Test cover the courses mentioned in the three secondary grades. Questions vary in the academic achievement sections in terms of the nature of their focus on cognitive levels. There are a number of questions that require understanding to answer and others requiring application, while a third type requires inference, and so on. Questions also cover the three High School grades in the following proportions: 20% for the first grade, 30% for the second grade, and 50 % for the third grade. The questions are divided in equal proportions for the following subjects: biology, chemistry, physics and mathematics.

2.2.4.2.2.2   The Academic Achievement Test for literary disciplines

The testing includes public High School graduates (the theoretical section and Quran Memorisation schools) and those who wish to attend colleges that are test-based. The academic literary Achievement Tests focus on general concepts in the following disciplines: Modern Islamic Culture, Monotheism, Jurisprudence, Grammar, Rhetoric and Criticism, Literature, and History and Geography. Academic achievement questions in the aforementioned courses include the three secondary grades (Theory/Quran Memorisation). Questions in the academic achievement department vary in terms of the nature of its focus on cognitive levels. Some of the questions require understanding, and others involve application, while others entail analysis. Questions cover three stages of the High School level with the following proportions: 20% for the first grade, 30% for the second grade, and 50% for the third grade of the theoretical departments.

All test questions (theoretical and scientific) are of the multiple-choice type (A, B, C, D). The student should highlight the letter corresponding to the correct answer for each question on the answer sheet using a pencil, which will be distributed in the exam room. As the questions include easy-to-answer maths questions, the use of a calculator, or any machine that acts as a calculator, is not allowed. Testing is expected to take about three hours, of which approximately one hour is dedicated to the procedures, instructions and to filling in the student information on the answer sheet, which should take place before the start of the test. The other two hours are divided into five sections, with each section lasting 25 minutes.

2.2.4.2.3   Results

After the end of the testing in all cities, the answers will be corrected and analysed; then, the correctness of information provided by the students will be centrally checked at the Centre in Riyadh. It is expected that the marking process will take about three weeks from the test date. As for the results, they will be announced after the marking process is finished.

Universities and colleges requiring a test receive the results electronically, and, thus, submission of such test results will not be required in paper form. The student has the right to inquire about the test results by:

1-   Calling a general contact number (00966920 033 555).

2-   Using the results' query page via the Internet:

http://www.QIYAS.org/QIYASResult/Pages/ResultsPage.aspx

2.2.5   A brief description of current admission standards at King Faisal University

Many universities in Saudi have typically relied on scores from High Schools for university admission. The National Centre for Evaluation and Measurement, QIYAS, was established in 2001 and published and applied the first test in 2002 to be used in all Saudi universities. The test was only required for male students until 2010, after which it included females. At KFU, the locus of the current study, only one study has addressed the admission standards at KFU and this was conducted by Alnajjar (2001). However, Alnajjar conducted his study before the launch of the National Centre for Evaluation, QIYAS.

This signifies the importance of conducting further research to investigate the admission standards and explore productivity for student success.

King Faisal University in Alahsa was founded in 1975 under royal decree by King Faisal before his death, but was opened in the era of his brother, Khalid Bin Abdul Aziz Al Saud. The university offers students the opportunity to pursue Higher Education and graduate studies in various academic fields, including science and specialised areas of knowledge and scientific research. King Faisal University has extended over an area of 480 hectares since its opening day in 1975 in the city of Alahsa overlooking the road leading to several GCC countries, such as Qatar, the UAE and Oman. In order to be admitted to this university, students are required to attain a High School diploma with a score that can vary from one

college to another. This is in addition to the test that must be conducted in some faculties and personal interviews held by a range of university faculties. These requirements may be consistent with the majority of the Saudi universities before 2002.

However, after the establishment of The National Centre for Assessment in Higher Education (QIYAS), there has been almost unanimous agreement among Saudi universities, including King Faisal University in Alahsa, on the general terms and conditions for admission to various colleges, including the College of Education at King Faisal University. Such terms and conditions include the following:

1- The applicant must have a High School diploma or its equivalent from outside the Kingdom of Saudi Arabia.

2- A period of five years will not have elapsed since the student has received a High School diploma or the equivalent.

3- The applicant must have passed the Abilities and Achievement Test hosted by the National Centre for Assessment in Higher Education.

4- The student must have good conduct.

5- The student should pass any test or personal interview proposed by the University Council.

6- The student should be medically cleared and provide proof of this from an accredited clinic or hospital.

7- The student should seek approval from his workplace if he/she is in employment.

8- Students will not be accepted if they have left any other university because of disciplinary or educational reasons.

9- Students should not hold a bachelor's degree.

10-   The student must not be attending another university inside or outside the Kingdom on a regular basis.

11-   The student does not meet any other conditions set out by the University Council and announced at the time of submission.

12-   Students will be accepted through competitive differentiation between the applicants based on the weighted percentage of High School, the Aptitude Test and the Achievement Test.

More specifically, the table below explains the terms and conditions that are required by the Education College at King Faisal University:

Table 2-3: the terms and conditions set out by the Education College at King Faisal University

Faculty	Type of secondary	Admission requirements
Education	All departments	High School diploma or equivalent with an overall percentage of 80% at least for both male and female students. Male students: the weighted percentage is 60% of the cumulative for High School and 40% of the Aptitude Test. Female students: the weighted percentage should be 50% of the cumulative percentage of the High School diploma, 20% of the Aptitude Test and 30% of the Achievement Test. There should be at least 85% weighted percentage to compete for an admission place.

2.2.6   *Vision 2030*

In the present era, Higher Education has undergone a series of changes as a result of globalisation. One of the most significant impacts on Higher Education is that it has become more open to other countries and people around the world. A range of concepts have emerged in Higher Education circles, including quality, marketing, internationalisation and diversity (Alzaki, 2017). These changes have had an impact on the method of work employed by Higher Education institutions and have been a major driving force behind the adoption of the concept of marketing services for universities, which has improved the quality of their offering and linked them directly to the needs of the labour market to encourage competition between universities. It has also resulted in improvements to their academic reputation and the creation of brands for Higher Education institutions in society (Alsaqri & Almuhaimed, 2017).

UNESCO considered internationalisation as one of the most significant challenges facing Higher Education, and this requires Higher Education institutions to be focused on openness, cultural exchange, scientific and research participation, as well as the activation of cooperation agreements with distinguished universities on the global stage, and adherence to international standards and practices adopted by advanced universities. As such, internationalisation has become a strategic choice for Higher Education institutions around the world to enhance their competitiveness and global status.

Hemsley-Brown and Oplatka)2010) noted that most universities have recently realised that they need to market their services, especially in the face of fierce competition from other universities in terms of the services they provide, which indicates that the concept of marketing has shifted from the business sector to the Higher Education sector. Similar to the

business world, universities are supposed to promote a unique brand and maintain an enviable reputation among their customers, who are, ultimately, the beneficiaries of their services (Alzaki, 2017).

Universities all over the world are competing with each other in the products they offer to consumers. One of the most important factors for these universities is to keep thriving and expanding the quality of their products and the suitability of these as regards consumers' requirements.

Yang (2003) believes that Higher Education as an investment in human capital has become a fundamental building block of educational policies in many countries and a crucial part in the development of citizens and countries alike.

Therefore, the Higher Education institutions in the Gulf are facing a great challenge in terms of achieving quality outcomes and increasing academic endeavours characterised by quality rather than quantity, in addition to dealing with the fierce competition among international universities, where investment policies in Higher Education have been adopted to provide compatibility with labour market requirements (Al-Ghanbousi, 2012). Moreover, these Higher Education institutions strive to provide high quality output as the sector's main objective. In other words, their priority is to send qualified students, who have just graduated, into the labour market to fill the gap in certain fields and transfer knowledge and skills to the work environment.

According to Khoury (2013), globalisation has had an impact on the Higher Education institutions in the Arabian Gulf in terms of demand for more ICT-related outputs, as well as monetary and creative skills. Thus, the improvement of Higher Education institutions has become mandatory in order to fulfil these requirements.

However, despite the effects of globalisation on Higher Education on a global scale and the positive action to meet its challenges and seize upon its opportunities, universities in Saudi

Arabia in general, and Education Colleges in particular, have not shown the response expected in order to capitalise on the opportunities that have arisen as a result of globalisation (Alameri, 2017).

Saudi Higher Education's lack of response to international changes imposed by globalisation in the field of Higher Education confirms the results of a number of studies (e.g. Al-Amri, 2014; Al-Kareani, 2010; and Al-Otaibi, 2006), which have all pointed to the failure of Saudi universities to benefit from research centres and international universities, as well as joint international research. There is also the issue of the lack of international academic partnerships for faculty teaching personnel in Saudi universities, in addition to the poor level of faculty teaching staff across faculties in terms of both developing their skills in the field of learning and learner assessment, and their application of educational practices in the delivery of modern subjects, which is reflected in a weak final output from the faculties of education.

Perhaps, it can be argued that the Saudi Arabian *Vision* 2030 has emerged as a potential factor in meeting the requirements of the era in various fields. The Saudi Arabian *Vision* 2030 is based primarily on three axes, namely a vibrant society, a thriving economy and an ambitious homeland. These axes complement each other in order to achieve the objectives of the *Vision*. The second element, the thriving economy, focuses on providing opportunities for all by building an educational system linked to the needs of the labour market. *Vision* 2030 is equally important because it is a well-defined road map that has been prepared to move the Kingdom of Saudi Arabia towards a prosperous future and ensure it is on the right track in terms of development. One of the most important objectives of *Vision* 2030 is concerned with education:

- Strategic Objective 1: To provide educational services to all students.

- Strategic Objective 2: To improve the preparation and ensure professional development of teachers.

- Strategic Objective 3: To enhance the educational environment that stimulates creativity and innovation.

- Strategic Objective 4: To develop curricula, teaching and assessment methods.

These two objectives are consistent with the aim of this thesis, which is focused on offering diverse means of learning assessment because of the positive impact this can have on the attainment of an accurate and fair result. At the same time, diversity in assessment and teaching methods may eventually contribute to identifying students who have creative skills and knowledge and would be in line with the requirements of the current labour market and Saudi Arabia's *Vision* 2030.

- Strategic Objective 5: To promote students' core values and skills.

- Strategic Objective 6: To enhance the capacity of education to meet the needs and expectations of the labour market.

- Strategic Objective 7: To diversify innovative sources of funding for the education sector.

- Strategic Objective 8: To increase the role of the private sector in education.

In other words, with regard to Higher Education, *Vision* 2030 aims to develop the entire educational system, starting with the systems and structures, creating labour market focused curricula, and, of particular relevance to the focus of this thesis, appropriate assessment methods, while ensuring that future generations are still connected to the traditions and religious values of society. This is highly important given the fact that 58% of the population is under 30 years of age in KSA. Therefore, to guarantee the future success of the country,

enormous investment in the young, who are the business leaders, economists and teachers of tomorrow, is required.

The '*Vision*' also took into consideration the importance of developing Higher Education in order to harmonise the educational output with the requirements of the labour market. It highlighted the need "to bridge the gap between the outputs of Higher Education and the requirements of the labour market" (Elmulthum & Elsayed, 2017, p.21). According to figures from the General Bureau of Statistics (2016), Saudi Arabia is still suffering from high unemployment rates, reaching about 12% of the total national workforce on average (6.3% for males and 34.8% for females). As shown in a report by Kirdar (2002), this high percentage may be attributed to the fact that many graduates of Arab universities were unable to find suitable jobs due to poor academic standards and the irrelevance of graduates' qualifications to the labour market. In addition, Al-Ghanbousi (2012) added that the admissions policies adopted in some Gulf universities limit the flow of the number of high-quality applicants aspiring to join Higher Education institutions from state education. Such policies include the standardised tests applied in most countries in the region that have been the subject of much criticism by members of the community, as well as in academia. One of the most important criticisms directed towards standardised tests is that they limit the fairness of the admission system and affect equal opportunities for all applicants (Al-Ghanbousi 2012; Ricketts 2010).

As mentioned previously, the *Vision* also aims to ensure that at least five Saudi universities are among the top 200 universities in the world by the end of 2030. Looking at the most important criteria of high performance at the universities, they would possibly include the quality of teaching staff, quality of outcomes, the level of student employability, student surveys (as in the National Student Survey), and research outputs (Abdul Qadir, 2018). Therefore, this study seeks to shape the context of Higher Education not only for KSA, but also with regards to what is happening in universities more generally.

In general, the *Vision* also includes increasing the participation of women in the labour market to 30%, perhaps by empowering Saudi women to become active players in the labour market and contribute as men do in building their Saudi homeland. The need to gauge the educational attainment of female students in public and Higher Education has become a priority and a necessity to achieve the aims of *Vision* 2030. This puts the spotlight on the educational achievement of female students in a conservative society. Therefore, the current research is important in terms of contributing to *Vision* 2030 by highlighting the educational achievement of female students and drawing the relevant comparisons with students in the Faculty of Education at King Faisal University.

In the field of education, activating the Kingdom of Saudi Arabia's *Vision* 2030 has become more procedural in all universities by establishing a dedicated department to achieve such a vision in each of the Saudi universities linked directly to the university Dean. These departments are the primary link between the university and the relevant authorities, overseeing the journey of transforming universities into independent universities. Among the most prominent tasks of these departments is to achieve the *Vision* 2030 in universities by presenting initiatives and projects that contribute to achieving such vision, while monitoring its implementation and submitting reports and outcomes. For instance, reaching a balance between university outcomes and the requirements of the labour market, as well as building specialised partnerships in the health and education field to raise the quality and efficiency of the services provided (KSU,2019).

While at King Faisal University, one of the most prominent goals of the establishment of an office to action *Vision* 2030, as follows:

- Directing the university's research activity to serve the national goals of the Kingdom of Saudi Arabia and the aspirations of the university's identity.

-        Utilising the university resources and capabilities for an exceptional university experience that capture the students' expectations.

-        Adopting best practices to enhance the university's institutional performance and its business governance efficiently and effectively.

-        Redesigning existing academic programs and creating academic programs consistent with *Vision* 2030 and the aspirations of the labour market (KFU, 2019, p. 27-30).

Perhaps the current research is geared towards the attainment of a number of national goals in education, especially in terms of developing the standards of admissions in the College of Education at King Faisal University. This research is also concerned about developing ways of how learners can be assessed, especially those linked to alternative assessment methods that have already been mentioned in this research and which are in keeping with the Saudi project *Vision* 2030.

Chapter 3   Literature review

3.1   Introduction

The previous chapter addressed the public education system in Saudi Arabia and the authorities supervising it. Then, it highlighted the types of High School, school subjects and details concerning arrangements for the assessment of pupils. Chapter two concluded by giving a brief description of the admissions system at King Faisal University in Saudi Arabia. The aim of this chapter is to provide the reader with an overview of the literature related to predictive validity. It will cover the definition of validity and its kinds. An extensive review of the literature on predicting future academic success is presented and will explain testing such as IELTS, TOEFL, SAT, GRE, GMAT and A-LEVELS. The chapter also discusses predictive validity in neighbouring countries such as the Gulf States and Jordan. This part of the thesis also reviews the literature on lecturers' perceptions about assessment. It also pays some attention to issues of assessment in school and in Higher Education.

3.2   Validity in assessment

In the evaluation of educational and psychological tests, validity has remained one of the most contested topics for scholars in linguistics and in education (Rumsey, 2013). According to Golafshani (2003) and Hassan (2006), validity is an important consideration in evaluating these tests because validity reflects the value of tools that are used in assessment and measurement. As test validity increases, so too increases confidence in the validity of results. Validity was first defined by Garrett (1937) as "The validity of a test is the fidelity with which it measures what it purports to measure (p.324, quoted in Rumsey, 2013: 12) A later definition some 63 years later by Bachman (2000) describes validity as a foundation upon which consideration of values, their uses and their effects are based. Over the years, how assessments are validated continue to be a great concern, but mechanisms used have

continued to evolve since the concept was first derived (Garrett, 1937). Messick (1994: 13) also describes validity as an 'integrated evaluative judgement of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores'. According to Moss (2007), unified validity as a conceptual theory is considered a reflection of scientific inquiry but in a general sense. As per Messick's theory, Keith (1998) agrees that validity is highly influential due to its ability to combine disparate contributors into a unified framework for developing validity arguments. Salkind (2013) also added that validity can be interpreted as a component of a tool that shows that the tool is doing what it should be doing. In order to bring the concept of unitary validity to light, it is vital to understand the issue being unified and the reasons why it is being done (Moss, 2007). There is some agreement that validity obtains when evidence on an issue is accumulated to a certain level and the theory being used supports a specific interpretation of the test scores in regard to that specific test (AERA, APA, and NCME, 2014; Alhajri, 2013; Brown and Coughlin, 2007; Hassan, 2006; Yeain, 2003).

3.2.1 Types of Validity

Validity is perceived in many ways in the literature that discusses issues concerning measurement and evaluation concepts. Some of the common types of validity include face validity, discriminant validity, empirical validity, Trustees' validity, and factorial validity, among many others. Generally, most of the references, which include the American Educational Research Association, The American Psychological Association and the National Council on Measurement in Education, tend to focus on three types of validity and these are content validity, criterion validity and construct validity (AERA, APA, and NCME, 2014). These types of validity are discussed in the following section.

3.2.1.1   Content Validity

Content validity is defined by Bollen (1989) as a qualitative type of validity in which a concept's domain is clarified and the analysis is used to determine whether the measures fully reflect or represent the domain. Content validity is basically a qualitative approach ensuring that a domain's indicator reflects the concept as defined by the analyst. The main weakness of the concept is its dependence on theoretical definition. Just like many other concepts in the social sciences, there is limited consensus on theoretical definitions, thus rendering the domain of the content ambiguous. Content validity burdens researchers to provide theoretical definitions that are acceptable to their peers and also to carefully select specific indicators that would fully cover relevant domains and dimensions (Bollen, 1989). Content validity also requires that, on top of a researcher providing a theoretical framework for the concept's definition that is acceptable to peers, the indicators selected for the concept should be comprehensive in covering the domain and its dimensions. Content validity therefore provides a means for researchers to qualitatively ensure that the indicators provided reflect the overall meaning of their concept as defined by the researcher (Drost, 2004). A good example can be derived from a researcher undertaking a test on a person's knowledge of elementary geography using pencil and paper testing. It will be important to ensure that the test represents the domain of elementary geography. The survey can test the individual's knowledge of elementary geography by asking general questions on the continents or by testing their advanced knowledge of geography on issues such as climate change in relation to continental topography. Content validity thus ensures that a researcher undertakes a test which represents the domain of the concept. Content validity can be assessed in a number of ways: by asking specific questions regarding the domain being tested, or by asking the opinion of experts in that field of study (Drost, 2004).

3.2.1.2   Criterion Validity

This refers to evidence derived from the relationship between constitutes of a measurement tool and its performance on other variables. This approach is divided into two common types: concurrent validity and predictive validity (DeVon et al., 2007).

3.2.1.2.1   Concurrent Validity

Concurrent validity is where a test being undertaken correlates with another test that assesses a similar construct at the same time (AERA, APA, 2014; Alhajri, 2013).

3.2.1.2.2   Predictive Validity

Of particular relevance to the topic of this thesis, predictive validity is the prognostic capacity assessment of exam scores in an effort to predict the future success of candidates. This can be likened to the assessment of the achievement of professional success during university education or how successfully a graduate performs professionally in the future (Panczyk & Gotlib, 2015). Various bodies including The American Educational Research Association (AERA), the American Psychological Association (APA) and the National Council on Measurement in Education (NCME) define predictive validity as an evidentiary concept based on relations to other variables, which include a test-criterion relationship. This indicates the accuracy of test data and how effectively they can predict criterion scores that can be obtained in the future. Statistical concepts that have used predictive validity include Pearson correlation coefficient, Linear, multiple regression, one-way ANOVA and Multivariate Analysis of Variance (MANOVA).

3.2.1.2.2.1   Advantages and Disadvantages of Predictive Validity

That this thesis has a particular focus upon predictive validity demands some detailed attention to this concept, which now follows. Formal, standardised examinations control

students' lives and the scores that they receive from these exams will define the opportunities that they have in order to access Higher Education and future success (Broadfoot, 1984). Predictive validity could be considered 'pesky' for its perseverance among various test stakeholders as a test quality which is eligible, despite the weaknesses that some researchers in the field of assessment criticise it as possessing. Since 2000 there has been an abundance of studies which have examined the predictive validity of various testing tools with regard to academic success (Rumsey, 2013).

In spite of being popular, there have been a number of serious criticisms of predictive validity. As highlighted by Brown (1990: 42), "prediction of future performance is always dubious. There is no certainty that motivation and opportunity to achieve will remain constant over time." As such, "assessments and records of what students have achieved cannot and should not be treated as predictive of what students can and will achieve [in the future]" (Sutton, 1991, as cited in Rumsey, 2013). On the other hand, according to McNamara (2001: 337), "our existing models of performance are inadequately articulated, and the relationship between performance and competence in testing remains obscure. In particular, the assumption of performance as a direct outcome of competence is problematic, as it ignores the complex social construction of test performance". Based on the viewpoints of these authors, it seems that while the grading tool is too greatly emphasised and the learning function less so, the entire social element of performance in testing does not seem to be well-grasped or given due consideration.

Nevertheless, recognising these criticisms does not conflict with the fact that these findings have been and will carry on being utilised in the selection standards. As stated by Paul Black (1998: 44), "it would be easy if all those who took a test were admitted to degree study, irrespective of their results, but this does not happen". There are two factors that may stop this from happening: first, it is clearly not realistic for organisations to acknowledge every single applicant. A particular assessment tool as part of the selection criteria will always be

required, and, second, there should be access for everyone to a degree programme, no matter what their results are, and this would also clearly do without requiring the assessment tool to start with. Finally, it is possible that the predictive validity of an assessment will help teachers in selecting what it is required to focus on or develop with the students in the future, in addition to the areas of learning which every single student has to pay attention to in the forthcoming term or weeks (Rumsey, 2013).

3.2.1.3 Construct Validity

The concept of construct validity is based on evidence integration that bears on the interpretation or meaning of a test score's result. This includes content- and criterion-related evidence that are modelled as part of construct validity (Messick, 1994). Construct validity involves generalisation of responses and results derived from various assessments that are related to other assessments. It involves indirect measurement that is theoretically based, including its assessment (Al-Dossary, 2001).

3.3 Predicting future academic success

3.3.1 Introduction

This section addresses the main measures for predicting future academic success via the Scholastic Aptitude Test (SAT), the International English Language Testing System (IELTS), the Test of English as a Foreign Language (TOEFL), the Graduate Record Examination (GRE) and the Graduate Management Admission (GMAT). This section explains each one of these exams and reviews and critiques papers which have investigated the predictive validity of these exams and academic success. In general, the average of predicting future academic success via exams taken and GPA at colleges was higher for female students than male students (Kim et al., 2003; Morgan, 1990; Young, 2001; Zwick,

2013). However, there are some studies that have found males to have higher performance in predicting future academic success than females (Sireci & Talento-Miller, 2006; Wainer & Steinberg, 1992). The literature review on gender differences in academic performance at different levels reveals varied outcomes. Higher Education studies show that females are often found to outperform males (Dayıoglu & Turut-Asik, 2007). Leonard and Jiang (1999) argued that female students have better study skills than their male counterparts. Wainer and Steinberg (1992) added that females receive higher grades than males because they work harder and attend lessons and classes more frequently.

According to some university websites in Gulf countries, the number of females that are accepted in Higher Education is higher than the number of males. For example, regarding the Gulf newspaper, 64% of students who were accepted in the Academic year 2014 into Emirati Higher Education were females. Also, at Kuwait University there were 37,234 students accepted in 2014, of which 71% of them were female (The Office of The Vice President for Planning, 2014). According to the Education Ministry in Saudi Arabia, there were 866,823 students accepted in all the colleges and universities in Saudi Arabia in 2010 and 477,370 of them were female. In addition, regarding the KFU website, the university accepted 19,754 students in 2010 for bachelor and diploma degrees. 12,050 of them were female. This current study includes females and the statistics show that female students outnumber male students in the Gulf countries and also in the universities of Saudi Arabia. This fact renders this study more interesting, and generates wide applications for the research because it will include a significant body of students – the majority, in fact – who have typically been neglected by academic enquiry.

3.3.2   SAT exam

The Scholastic Aptitude Test (SAT) is the administration standards form that was designed in 1926 (Brown et al., 2002). The SAT is a globally recognised college admission test that

has enabled institutions to access information that analyses how well individuals are performing and if they can effectively apply their knowledge. The platform tests an individual's reading, writing and maths skills. These are common subjects taught in most High School syllabi. The SAT is mostly taken by students during their junior or senior year of High School. The exams are used by many colleges and universities in the U.S.A to make admission decisions (see www.sat.collegeboard.org).

The SAT tests general skills that students acquire in school. Reading skills are tested by reading passages and completing sentences, while the writing test involves essay writing and answering questions of a multiple-choice format. This involves the identification of errors in sentences and the use of grammar in sentences. Maths tests involve arithmetic calculations, algebra, geometry and questions with statistics and probability. Table 3.1 summarises the skills that the SAT examines, the duration of the test and the types of questions.

Table 3-1: Summary of the skills that the SAT examines

Skills and Points	Time	Question Number and Types
Reading **200-800 points**	70 Minutes One 20-minute section. Two 25-minute sections	(48 Passage-based reading) Tests student's comprehension of what is stated in or implied by the passage.
		(19 Sentence completion questions) Tests student's vocabulary and their understanding of sentence structure.
Math **200-800 points**	70 Minutes One 20-minute section Two 25-minute sections	(44 Multiple-choice) The students are asked to solve a problem and pick the best choice offered.
		(10 Student-produced responses) A student is not given answer choices. He/she must solve the problem and "grid in" their answers.
Writing **200-800 points**	60 Minutes One 10-minute section Two 25-minute sections	(25 Improving Sentences) Tests a student's ability to correct faults in usage and sentence structure and recognize effective sentences that follow the conventions of Standard Written English.
		(18 Identifying sentence errors) Tests a student's ability to recognize faults in usage, and recognize effective sentences that follow the conventions of Standard Written English.
		(6 Improving Paragraphs) Tests a student's ability to revise sentences in the context of a paragraph or the entire essay, organize and develop paragraphs in a coherent and logical manner, and apply the conventions of Standard Written English.
		(1 Essay) The students are asked to present and support a point of view on a specific issue. Because the students have only 25 minutes, their essays are not expected to be polished - it is meant to be a first draft.

3.3.3 International English Language Testing System (IELTS)

This is a proficiency test in English language given to people who wish to pursue Higher Education or seek global migration. Over the last two years, over two million tests have been taken (British Council website, 2015). IELTS exams are designed jointly by the British Council, the IDP (International Development Program), and Australia and Cambridge ESOL (English for Speakers of Other Languages). IELTS have more than one thousand locations and test centres in more than 140 countries (see www.ielts.org.) IELTS have three types of exam, and these include:

3.3.3.1 Life skills IELTS

This test was introduced in April, 2015. For one to obtain a United Kingdom Visa (UKVI), one must attain 4.0 or above on the IELTS Speaking and Listening Exams. The test result acts as the proof of English level for settlement and a family visa (British Council website, 2015).

3.3.3.2 General training IELTS

People who migrate to other English-speaking countries like New Zealand, Canada and Australia for work experience, to study High School or to undertake preparation courses sit this exam. This exam is based on the basic skills of listening, speaking, reading and writing (British Council website, 2015).

3.3.3.3 Academic IELTS

This is a requirement for people who apply to study in English-speaking institutions of higher learning or universities or professional registration. Admission to such institutions for either undergraduate or postgraduate courses depends on the Academic IELTS results (British Council website, 2015). This type of IELTS exam is also divided into four skills: speaking, reading, writing and listening. This takes 161-165 minutes. The sections of the IELTS academic test are now addressed in more detail.

3.3.3.3.1 Reading (60 minutes)

This part consists of 40 questions. There are a variety of questions, including identifying information as true, false or not given, or multiple-choice questions. Questions can also involve identifying the writer's views as either yes, no or not given. Additionally, matching of information, features, headings or sentence headings, the completion of sentences, notes, tables, flowcharts, diagram labels, and short answer questions are also tested (UCLES, 2013). This part includes three sections, totalling 2150-2750 words. Each of these sections contains an authentic long text. These texts range from descriptive texts to analytical and discursive ones. They are accessible to any candidate who wants to join undergraduate or postgraduate courses or who is seeking to be registered in a professional organisation. These texts are sourced from journals, newspapers and magazines or books. They are usually on general academic topics of interest. They are not originally specialised for any audience. These texts sometimes contain non-verbal materials such as graphs, diagrams or other illustrations (UCLES, 2013).

3.3.3.3.2 Writing (60 minutes)

There are two tasks. Task 1 requires one to write 150 words, while task 2 requires 250 words. These tasks relate to issues of general interest to the candidates. These issues are easy to understand and are suitable for candidates seeking to join institutions of higher learning for undergraduate or postgraduate courses, or those who seek professional registration. In task 1, candidates are asked to explain or describe graphs, diagrams tables or charts using their own words. Sometimes candidates are asked to explain how something works, how an event occurs or how an object is produced. Candidates may also be asked to explain the stages of a given process. However, in task 2, the candidates consider an argument and write an essay. They may also write one from a required point of view. It should be noted that all responses by candidates to these to tasks must always be in a formal style (UCLES, 2013).

3.3.3.3.3 Speaking (11 to 14 minutes)

The candidates meet with the examiner face to face and they are interviewed verbally. The conversation is usually recorded. The conversation has three parts: Part 1 is the introduction and interview (4-5 minutes). This part constitutes the introduction of both the examiner and the candidate to each other. The candidate should also confirm their identity. This is followed by general questions from the examiner to the interviewee on familiar topics such as family, place of residence, work, and interests. Part 2 is the individual component (3-4 minutes). The candidate is given a task card. This card asks the candidate to speak on a given subject. The card also contains points which the candidate should cover while tackling the subject. The candidate is then given one minute to prepare, and they are permitted to make some notes. The next 1-2 minutes are for the candidate to speak about the subject, after which the examiner asks several questions related to the same subject as the question card (UCLES, 2013). Part 3 is the two-way discussion (4-5 minutes). The examiner continues to ask more questions related to the subject and the candidate is allowed to expand on the important points (UCLES, 2013).

3.3.3.3.4 Listening (30 minutes)

This part is divided into four sections and 40 questions. There are a variety of questions including the labelling of maps, plans or diagrams. There can be multiple-choice questions. Completion of sentences, notes tables, flowcharts, diagram labels and summaries, as well as short answer questions are also tested (UCLES, 2013).

Section 1: This involves a conversation between people in a daily social context (for example, two people in an accommodation agency). Section 2: This involves a monologue. It is also based on a daily social context (for example, a conversation about the arrangements for meals in hotels). Section 3: This involves conversation among people (up to four) in a training context (for example, a group of four students). Section 4 is a prologue on a subject

that is academic. Every section is heard once and there are different native speaker voices employed. In IELTS, candidates are not graded pass or fail, but rather on their performance based on a scale of 1 to 9 for each test. As mentioned, the tests include speaking, reading, writing, and listening. These test scores give the Band Score. The level of competence for the students determines the band score. The band scores are reported in either whole or half band scores, e.g., 6.5, 7.0, 7.5, and 8.0. IELTS has endorsed 2 years as the period of validity for the scores. The meaning of the band scores and the guidance on setting appropriate band score requirements are shown in tables 3-2 and 3-3.

Table 3-2: Explanations of IELTS scores.   (IELTS.org)

BAND	DESCRIPTION
9	Expert user
8	Very Good User (Fully Operational)
7	Good User (Operational Command)
6	Competent User (Generally Effective Command)
5	Modest User (Partial Command)
4	Limited User (Basic Competence Limited to familiar Places)
3	Extremely Limited User
2	Intermittent User
1	Non-User
0	Did not attempt the test

Table 3-3: Provides guidance on setting appropriate band score requirements. (ielts.org)

Band	Linguistically demanding academic courses (Medicine, Law, Linguistics, Journalism, Library Studies)	Linguistically less demanding academic courses (Agriculture, Pure Mathematics, Technology, Computer – based work, Telecommunications)	Linguistically demanding training courses (Air Traffic Control, Engineering, Pure/Applied Sciences, Industrial Safety)	Linguistically less demanding training courses (Animal Husbandry, Catering, Fire Services)
9.0-7.5	Acceptable	Acceptable	Acceptable	Acceptable
7.0	Probably acceptable	Acceptable	Acceptable	Acceptable
6.5	English study needed	Probably acceptable	Acceptable	Acceptable
6.0	English study needed	English study needed	Probably acceptable	Acceptable
5.5	English study needed	English study needed	English study needed	Probably acceptable

According to table (3-3), English-competence for all academic programs is applicable when a student has a score of 7.0 or above on the IELTS test. The programs with the lowest linguistic requirements require a score of 5.5 for admission.

3.3.4   GRE exam

The majority of schools in the United States require that the Graduate Record Examination (GRE) be taken by prospective scholars. The GRE exam is a regular test for admissions (ets.org/gre). This exam is facilitated by the Educational testing Service (ETS) which was established in 1949 (Young et al., 2014). The main aim of the GRE exam, according to the ETS, is evaluating the verbal reasoning of scholars, critical thinking, reasoning that is quantitative, and writing that is analytical. These skills have been key for a very long period and are not linked with any field of study. Parametric testing centres offer a GRE exam that is computer-based. A revised GRE general test was issued in 2011. This test measures critical thinking, analytical reasoning, reasoning that is quantitative, and verbal reasoning. These skills have grown for a couple of years and are not linked whatsoever with any other field of study. This revised test includes types of questions that display the kind of thinking one will have and the necessary skills to enable one to excel at business school (ETS, 2012).

Table 3-4: The GRE exam part, the number of questions and the time given (Educational Testing Services, 2015)

Measure	No of Questions	Time	Score scale	
			Before 2011	**After 2011**
Analytical Writing (2sections)	Section 1 Analyse an Issue and Section 2 Analyse an Argument	30 mins per section	0–6, in half-point increments	0–6, in half point increments
Verbal Reasoning (2 sections)	25 questions per section	35 mins per section	200–800, in 10-point increments	130–170, in 1-point increments
Quantitative Reasoning (2 sections)	25 questions per section	40 mins per section	200–800, in 10-point increments	130–170, in 1-point increments

According to the GRE website, GRE test scores are valid for five years following the year of testing.

3.3.5   GMAT

The Graduate Management Admission Test (GMAT) website is open all year and enables users to exhibit skills that corporate recruiters and admission officers for business schools are interested in. The primary need for the GMAT website is to aid admissions officers for graduates. GMAT came into use for the first time in 1954 (Sireci & Talento-miller, 2006). The GMAT website has reported that 4800 programmes for graduate management use their scores in facilitating admissions and more than 300,000 individuals have taken the GMAT test in previous years. Quantitative and verbal evaluations, when combined, form the GMAT aggregate score (Mba.com). The four parts that make up the GMAT test include verbal, integrated reasoning, assessment of writing, and a quantitative section. The duration for these sections is 3 hours and 30 minutes.

Table 3-5: The GMAT exam, the number and type of questions and the time.

GMAT Test Section	Questions	Questions Types	Timing
Analytical Writing Assessment	1 Topic	Analysis of Argument	30 minutes
Integrated Reasoning	12 Questions	Multi-Source Reasoning, Graphics Interpretation, Two-Part Analysis and Table Analysis	30 minutes
Quantitative	37 Questions	Data Sufficiency and Problem Solving	75 minutes
Verbal	41 Questions	Reading Comprehension, Critical Reasoning and Sentence Correction	75 minutes
Total Exam Time	____	____	**3hrs, 30 minutes**

3.3.6   TOEFL exam

One of the two most important tests for English-language proficiency is the Test of English as a Foreign Language (TOEFL) exam, the IELTS being the other major assessment

(ets.org/TOEFL). The Educational Testing Services (ETS) is the body in charge of preparing these tests. TOEFL is a test that is standardised for the ability to communicate in the English language for those whose native language is not English and those who are to enrol at USA campuses. The ETS website states that TOEFL is the most used exam for testing English as a foreign language in the whole world. TOEFL was first overseen in 34 countries in the year 1964 (Yan, 1995). This test has received recognition from 9000 universities, agencies, and colleges in over 130 nations, including the US, Canada, the U.K., and Australia (Vu & Vu, 2013). It is permitted for TOEFL exams to be provided as hard copies in areas around the globe. The TOEFL hardcopies are known as TOEFL PBT. This test is overseen in paper format, with the expected duration of completion being 3 hours and 30 minutes. The TW test is part of the PBT test and takes 30 minutes to complete. This paper is a requirement for everyone taking the paper format exam. This makes the total completion time for the PBT test 4 hours. The TOEFL PBT test has three major sections and the additional TWE writing test and these are all compulsory. The requirements are that one should answer the question in each section and write an essay to qualify for a score. TOEFL PBT scores typically range from zero to six hundred and seventy-seven. The comparison between TOEFL IBT, TOEFL PBT and IELTS test scores is shown in table (3.7) further down. TOEFL IBT is the most recent version of the TOEFL test and is delivered online. TOEFL IBT came into operation in September 2005 in the United States and between 2005 and 2006 in other countries (Service, 2015). This test focuses on integrated skills and, due to this, it provides more information about the learner's communicative ability and their readiness to work academically. The launch of TOEFL IBT in the United States has seen it being expanded to other regions. 96 percent of those examined take the TOEFL IBT test and the numbers of test takers keep growing. This means that access is expanding. Since the TOEFL IBT test measures the four basic communication skills, it has been preferred by a number of higher learning institutions around the globe. The TOEFL IBT test makes it possible for students to

facilitate their admissions to and exits from higher learning institutions, apply for visas, keep track of their academic progress, and apply for scholarships and candidate certification. Individual institutions and educational agencies set their test requirements, and, for this reason, there is no failing or passing in TOEFL tests. A combination of all the communication skills is desired in a TOEFL test. All those examined need to perform tests that involve all communicative skills, and more details are given in table 3.6 below.

Table 3-6: The TOEFL IBT sections, time, questions and tasks.   www.etsglobal.org

Section	Time Limit	Questions	Tasks
Reading	60-80 Minutes	36-56 Questions	Read 3 or 4 passages from academic texts and answer questions.
Listening	60-90 Minutes	34-51Questions	Listen to lectures, classroom discussions and conversations, and then answer questions.
Break	10 Minutes	-----------	-----------
Speaking	20 Minutes	6 Tasks	Express an opinion on a familiar topic; speak based on reading and listening tasks.
Writing	50 Minutes	2 Tasks	Write essay responses based on reading and listening tasks; support an opinion in writing.

Table 3-7: Comparison between TOEFL IBT scores, and the corresponding TOEFL PBT and IELTS band

TOEFL IBT Score	TOEFL PBT Score	IELTS Band
0–31	0–397	0–4
32–34	400-417	4.5
35–45	420-450	5
46–59	453-497	5.5
60–78	500-547	6
79–93	550-583	6.5
94–101	587-607	7
102–109	610-633	7.5
110-114	637-647	8
115-117	650-663	8.5
118-120	667-676	9

3.4   Studies on Predictive validity

A series of sections now follow which review the literature on predictive validity for the range of tests and assessment items mentioned above.

3.4.1   IELTS

IELTS is considered by some to be the world's best-known Aptitude Test for English language in Higher Education (British Council Website, 2015). Several articles and empirical studies have examined IELTS. This section highlights a number of such studies in an attempt to explain the validity of IELTS by a comparison of success and achievement in universities or colleges. Of the prognostic legitimacy research work, several studies have argued for the absence or weakness of the statistical connection between IELTS and students' academic performance. As has been stated, predictive validity studies examine how well students will do in the future (Alhajri, 2013). Some studies have investigated the predictive validity of IELTS. Table 3.8 below summarises some studies that have investigated the relationship between IELTS and academic success or GPA.

Table 3-8: Summarises studies that examined the relationship between IELTS & GPA

Study	Place	Sample	Kind of correlation	Correlation
Cotton & Conrow (1998)	Australia	33 undergraduate and postgraduate students	IELTS and GPA	0.24-*
			IELTS and staff rating	0.15*
			IELTS and students' self-assessment	0.28*
Dooey & Oliver (2002)	Australia	89 in Business, Science and Engineering	IELTS & academic Success	Non-significant
Kerstjen & Nery (2000)	Australia	113	IELTS and GPA	Non-significant
Yen & Kuzma (2009)	Britain	61 Chinese Business school students	IELTS and GPA	0.46**
Al-Malki (2014)	Oman	77 students and teachers	IELTS and CGPA	0.486**
			IELTS and TPC	0.194

*Correlation is significant at the 0.05 level. ** Correlation is significant at the 0.01 level.

The Table above reveals that only two of the studies reviewed found no statistically significant predictor of academic success (Dooey & Oliver, 2002; Kerstjens & Nery, 2000). The remaining studies found a statistically significant relationship between IELTS and Grade Point Average (Al-Malki, 2014; Cotton, & Conrow, 1998; Yen & Kuzma, 2009).

A mixed-method study to assess the level of IELTS prediction regarding the difficulty experienced by international students and the connection between the students' outcomes and IELTS was conducted by Cotton and Conrow (1998) at the University of Tasmania. First, questionnaires were employed in the collection of data for a period of one year, with a total of 33 participants. A further 23 students were interviewed, while a total of 34 surveys were returned by the participatory academic staff, and, eventually, two English support lecturers and two international student advisors. Data correlation was computed for the investigated students' IELTS scores and the three academic accomplishment measures: students' personal ratings of excellence, as well as the academic staff rating of students' achievement, and grade point average. In the main, there was no positive correlation observed. Nevertheless, the writing and reading subtest outcomes correlated at 0.34 and 0.36 respectively, being the students' achievement as rated by the staff. Additionally, a correlation of 0.39 and 0.46 was realised for students' personal approximates of academic achievements in the subsequent semester. As for the connection between the students' difficulties reported regarding language performance and the score of IELTS, no positive correlation was realised. The qualitative data spoke to the complexities in language as one of a number of factors having extreme impacts on the students' excellence. Many important prevailing variables were momentarily explored, such as motivation, the level of English language teaching offered, welfare, and cultural adjustment. Additionally, staff and students provided their views concerning IELTS, which was, in general, observed as being a reasonable test as stipulated by Cotton and Conrow (1998).

Identical findings were reported by Dooey and Oliver (2002) in their research on IELTS' analytical validity. Their study's objective was to access the analytical validity of an IELTS test perceived as a replica of potential educational accomplishment among 89 students taking science, business and engineering courses at Curtin University. A sample of 15 participants,

each bearing a distinctive nationality as well as 13 first languages, was considered. The majority of participants hailed from South-East Asia, particularly Indonesia and Malaysia. The sample, however, targeted a number of first-year undergraduates from non-English speaking regions who received admission due to their IELTS performance. There was no noteworthy connection found between IELTS outcome and the students' achievement within their disciplines. The study module served as the greatest evaluator of students' performances and it was the only solitary subtest of the four micro-abilities to attain an important correlation (Dooey & Oliver, 2002).

A study conducted by Kerstjens and Nery (2000) assessed the connection between the IELTS test and students' performance. Explicitly, the study's objective was to seek the level to which IELTS tests predict consequent student achievement, in addition to language complexities, for international students at a higher learning institution in Australia. Kerstjens and Nery equally endeavoured to assess if any of the individual examinations of reading, listening, speaking and writing was crucial to academic achievement. The assessment was conducted in the following ways: via statistical connection between the IELTS scores of the students and Grade Point Average, staff interviews, and student questionnaires. A correlation of 113 first-year international students' (from the Higher Education sectors of an Australian university business faculty and TAFE) IELTS scores against their initial-semester GPA was conducted. On the whole, considerable correspondences were found between the writing and reading tests and GPA (0.204, 0.262 respectively). Upon examining the scores for TAFE and Higher Education independently, only the Higher Education reading score remained significant. Although none of the correspondences seemed noteworthy in the TAFE group, the level of the correlation between GPA and the writing test (0.194) was identical to that of the whole sample, which seemed statistically important. Regression assessment realised a small-to-medium extrapolative impact of academic excellence from the IELTS score for the

whole sample, as well as the Higher Education board representing 8.4% and 9.1%, in that order, of the disparity in academic excellence. The single highest determiner of academic performance was the reading test from the Higher Education group and the whole sample. As stipulated by Kerstjens and Nery (2000), IELTS was deemed to be a non-significant determiner of the group's academic performance via TAFE.

From their study, Yen and Kuzma (2009) examined the correlation utilising a homogenous group of HND (Higher National Diploma) Chinese students who had been schooled at the Worcester Business School and specialised in business and management studies during the 2007/2008 and 2008/20009 academic years. While the Chinese student group was characterised with age similarity as well as schooling in business subjects within identical schools, this study seems designed to assess the likelihood of a considerable correspondence between students' performance and such IELTS findings. Yen and Kuzma (2009) reported a considerable correlation of .46 between the general IELTS result and GPA of the initial semester in addition to a lower connection (at 0.26) between the general IELTS result and the GPA for the second semester. Additionally, the study uncovered an important correspondence between the GPA for the first-year students' first semester and their writing, listening, and reading scores. The solitary skill not concurrent with the GPA of the first-year students within their initial semester was speaking. The study also reported that the students' GPA for the second semester as well as their IELTS listening scores were correlated (Yen & Kuzma, 2009).

More in line with the context of this thesis, another study conducted by Al-Malki (2014) assessed the connection between Omani English teacher apprentices and IELTS testing by employing a quantitative approach in gathering data. Ninety-four participants (Omani apprentice teachers) facilitated the data collected, which entailed their College Grade Point Average (CGPA) and expertise, as well as their teaching experience. The researcher

identified a mild correspondence between IELTS and CGPA, 0.486, which was statistically important. Bands of 7.0, 6.5, and 6.0 were obtained by all 77 teachers, which indicated the teachers' proficiency in teaching English subjects. However, the correlation between the IELTS test and Total Professional Competencies (TPC) was 0.194, which points to a weak relationship between IELTS and Total Professional Competencies (Al-Malki, 2014).

3.4.2   Critique of previous studies

In general, we can see that the results that come from the aforementioned studies are not similar. Some studies revealed a significant correlation between IELTS and GPA (Cotton & Conrow, 1998; Yen & Kuzma, 2009) and between the IELTS result and CGPA (Al-Malki, 2014). However, other studies found that a non-significant correlation existed between IELTS and GPA, or academic success (Dooey & Oliver, 2002; Kerstjens & Nery, 2000). Furthermore these studies have targeted many nationalities, though mainly South- East Asian students (Cotton & Conrow, 1998; Dooey & Oliver, 2002; Yen & Kuzma, 2009) and only one study was undertaken in an Arab country (Al-Malki, 2014). The studies involved different faculties, i.e., Business, Science, Engineering, Law, Humanities and English Language (Al-Malki, 2014; Cotton & Conrow, 1998; Dooey & Oliver, 2002). Al-Malki (2014), in his study, attempted to find a relationship between IELTS and teaching proficiency of teachers. He found a weak correlation between them. This could suggest that the IELTS test is a measurement of English language skills but not the teaching proficiency of the teachers.

From my review, I did not find any study in my region that investigated the predictive validity between IELTS and GPA, or academic success. This may be because some universities tend to adopt a local exam that is prepared by an English department. Furthermore, surprisingly, no study has been found investigating the predictive validity between IELTS and academic success in Colleges or Departments of Education at

universities. Finally, none of these studies used interviews to follow up the investigations. Such an approach can be employed to explain and interpret the factors involved and confirm preliminary analyses. Therefore, interviews would seem to be advisable in studies that investigate the relationship between pre-admission stats and GPA scores. 'Interviews are a popular and widely used means of collecting qualitative data' (Burns 1999, p. 118) and the investigator in this thesis wants to have first-hand information directly from some knowledgeable informants (Zohrabi, 2013). Through interviews, the researcher intends "to obtain a special kind of information" (Merriam, 1998, p. 71) and better interpret robust data that cannot be explained without the input of the participants (Zohrabi, 2013).

3.4.3   TOEFL

The Test of English as a Foreign Language (TOEFL) is intended to measure the English language skills of people whose mother tongue is not English. Some universities and colleges in more than 130 countries use TOEFL as a measure of academic success (Service, 2015). Many studies have discussed the predictive validity of TOEFL in different contexts. Table 3.9 summarises the results of some of these studies, which are then discussed in more detail.

Table 3-9: Summary of studies discussing the predictive validity of TOEFL in different contexts

Study	Place	Sample	Kind of Correlation	Correlation
Vinke & Jochems 1993	Netherlands	90 students from Indonesia	The TOEFL and GPA	TOEFL score < 450 = 0.09** TOEFL score > 450 = 0.50**
Hill, Storch, & Lynch 1999	Australia	55 students who took IELTS and/or TOEFL	The TOEFL, IELTS and academic success	IELTS and achievement =0.54 not significant TOEFL and achievement = 0.287 not significant
Wait & Gressel 2009	UAE	6516 students	TOEFL and GPA	TOEFL and GPA 0.339*
Cho & Bridgeman 2012	USA	2,594 students	TOEFL and GPA	Graduates 0.16 * Undergraduates 0.18*
Vu & Vu 2013	USA	610 students	TOEFL and GPA	0.07 not significant,

*Correlation is significant at the 0.05 level.   ** Correlation is significant at the 0.01 level

Table 3.9 reveals that only one study found no statistical significance as a predictor of academic success between the TOEFL test and academic success at university (see Vu & Vu, 2013). However, other studies found a correlation between TOEFL and university success (Cho & Bridgeman, 2012; Hill et al., 1999; Vinke & Jochems, 1993; Wait & Gressel, 2009).

Vinke and Jochems (1993) correlated the scores of 90 students from Indonesia who had taken an 11-month course in Sanitary Engineering at Delft University in the Netherlands between 1987 and 1991. They attempted to find a relationship between TOEFL and Grade Point Average scores. They divided the sample into two subgroups depending on the TOEFL scores: the first group was made up of students who had scores below the mean of 450 in the TOEFL exam. The second group was made up of the students who scored 450 or above in the TOEFL exam. Vinke and Jochems also divided the sample into two categories depending on age: the first group comprised of students who were less than 33 years old, and the second group was made up of students who were older than 33. They found that the strength of the correlation coefficient depended on the range of TOEFL scores and the age of students. The TOEFL scores which were lower than 450 points showed a weak correlation coefficient ($r = 0.09$), but the correlation coefficient strength ($r = 0.5$) started with scores over 450 points. The correlation was higher for the students who were younger than thirty-three years of age ($r = 0.64$ age <33 and $r = 0.38$ age>33).

Hill, Storch & Lynch (1999) examined the relationship between IELTS, TOEFL and academic success for international non-English speaking background students at the University of Melbourne in Australia. The study included 55 students who had taken IELTS and/or TOEFL. The relationship between achievement and IELTS results were found to be, on average, strong ($r = .540$). They found a weak linear relationship between IELTS and academic achievement by using a scatterplot for GPA. The correlation coefficient between

TOEFL results and GPA was (r = .287). An investigation of the scatterplot for GPA with TOEFL results showed a comparatively curvilinear relationship between TOEFL and GPA.

Wait and Gressel (2009) explored the relationship between TOEFL scores and academic success for international Engineering students. The sample comprised 6516 students from 85 different nationalities who enrolled at an American-style University located in the United Arab Emirates. 75.4% of the sample was made up of Middle Eastern/North African nationals and 24.6% came from other countries, for example, the UK, the USA, and Canada. By using linear regression models, the study found that the TOEFL with GPA coefficient is positive for all four colleges (Business, Arts & Science, Architecture, and Design and Engineering) and within all engineering majors (Chemical Engineering, Computer Science, Electrical Engineering, Mechanical Engineering, Computer Engineering and Civil Engineering). The regression coefficients were statistically significant at the level of = 0.05 for all colleges and majors. However, the level of significance was low in computer engineering, 0.061 and electrical engineering, 0.080. The highest regression coefficients between the colleges for GPA with TOEFL was in the Arts and Sciences College (TOEFL coefficient 0.00369).

Cho and Bridgeman (2012) examined the relationship between TOEFL Internet-Based Test (IBT) scores and academic performance. The sample comprised 2594 students, 1850 graduate and 744 undergraduate students chosen from 10 universities in the United States of America. The result from this study was that there exists a small correlation coefficient within TOEFL IBT and GPA. The average of correlation coefficients for TOEFL IBT for the group of undergraduate participants was (r = .18) and (r = .16) for graduates. This explained 3% of the variance of GPA for undergraduates and graduates. Another study that explored the predictive validity of TOEFL was completed by Vu and Vu (2013) and centred on international graduate students. The data included 464 students at a Midwestern public university in the U.S.A. The sample was chosen by a survey. The correlation between

TOEFL and GPA was 0.07, which means there was no correlation between TOEFL scores and GPA.

In critiquing this literature, there have been a number of studies that have addressed the relationship between TOEFL and academic success. Some studies found a correlation between TOEFL scores and academic success or GPA at university (Cho & Bridgeman, 2012; Vinke & Jochems, 1993; Wait & Gressel, 2009). However, there are some studies which could not capture any correlations, or, in some cases, only weak correlations between TOEFL and academic success (Hill et al., 1999; Vu & Vu, 2013). However, the common trend among these studies was that they all used a quantitative method, with no further follow-up investigation of factors such as interviewing participants or stakeholders about potential factors. Besides, the samples that were explored in the previous studies are either students speaking English as a second language in an English-speaking country, or students at English speaking universities in an English second language country. Finally, it can be seen from the summary table that some studies have low sample sizes.

3.4.4   SAT

The Scholastic Aptitude Test (SAT) is the main admission requirement for university study since it was first approved in 1926 (M. Brown et al., 2002). According to the SAT website, the exam is used by many colleges and universities in the United States to make admission decisions. Several studies and reports have discussed the predictive validity of the SAT as regards the academic performance of university students with a range of outcomes. The following Table (3.10) summarises the results of some of these studies.

Table 3-10: Summary of studies that examined SAT and the GPA.

Study	Place	Sample	Kind of Correlation	Correlation
Morgan 1990	USA	278,074 students	The SAT and FGPA by gender	HS-GPA and SAT with FGPA for female: 0.56, 0.54 and 0.53 HS-GPA and SAT with FGPA for male:0.53, 0.49, and 0.48
			The SAT and FGPA by major	HS-GPA and SAT with FGPA for Business: 0.57, 0.55 and 0.52 HS-GPA and SAT with FGPA for pre- professional: 0.56, 0.55 and 0.53
			The SAT and FGPA by ethical	HS-GPA and SAT with FGPA for white: 0.54, 0.52 and 0.50 HS-GPA and SAT with FGPA for Asian: 0.51, 0.47 and 0.48
Kobrin et al. 2008	USA	151,316 students	SAT and GPA	HSGPA with 1st-year GPA = .36 HSGPA & the three SAT subtest with 1st-year GPA =.46
Geiser and Studley 2010	USA	77,893 students	The SATI & The SATII UCGPA	UCGPA and SATI .07,.01, .02 and .11 (not significant)
				UCGPA and SATII = .21, .24, .26 and .22
Shaw et.al 2012	USA	39,440 students	SAT and cGPA	SAT and cGPA: math and statistic/physical sciences = .64 Education = 0.59

Table 3.10 shows that only one study found no statistically significant predictor of academic success between the Scholastic Aptitude Test (SAT) and freshman Grade Point Average (FGPA) (see Geiser & Studley, 2010). All other studies found a statistically significant relationship between academic success and the SAT (Kobrin et al., 2008; Morgan, 1990; Shaw et al., 2012).

Morgan (1990) examined predictive validity in categories of college students. Morgn had useful summaries of studies conducted since 1976 until 1980. His sample study comprised 278,074 students in the United States. The study sample included 222 colleges that conducted 778 different validity studies over the 10-year period. For each of the 778 analyses included in this study, the mean, standard deviation, and correlations among the three predictor variables and F-GPA were extracted from the VSS data base. Morgan examined

the predictive validity of the Scholastic Aptitude Test (SAT), the Test of Standard Written English (TSWE), and Achievement Tests which are prepared by the College Board within groups classified by sex, race, and college in 1978, 1981, and 1985. According to the results, females have higher correlation estimates than males. The SAT Mathematical (SAT-M) was the best predictor of students' (male and female) Grade Point Average (0.53, 0.50 and 0.49). The highest relationship between High School Grade Point Average (HS-GPA) and SAT with FGPA based on ethnicity (Asian Americans, black Americans, Hispanics, and whites) were white: 0.54, 0.52 and 0.50, followed by Asian students. SAT scores were better predictors of FGPA for black Americans (0.49, 0.42 and 0.41). The highest relationship between HS-GPA and SAT with FGPA by major (Business, pre-professional Liberal Arts, Technical and Pre-Professional) was Business: 0.57, 0.55 and 0.52, followed by Pre-professional..

An attempt to investigate the ability of the SAT in respect of predicting the GPA for the first year of college was made by Kobrin et al. (2008). They selected 151,316 students as a sample. As implied by this study, the first year Grade Point Average (FYGPA) had a correlation prior to correction with High School Grade Point Average (HSGPA) of 0.36 and the FYGPA had numerous correlations with the three SAT exams (writing, critical reading and mathematics) of 0.35. The results, after being modified by Kobrin et al. for correlations, were 0.54 and 0.53 respectively. In the context of each SAT test section, SAT-Writing had the highest pre- and post-correlation corrections that were 0.33 and 0.51 respectively.

Geiser and Studley (2010) examined the relationship between freshman grades and SAT scores based on accumulated data from 77,893 students who were newly admitted to the University at California (UC) during the period from fall 1996 to fall 1999. The study indicated that, as a single predictor, SAT II scores were the best for University of California Grade Point Average (UCGPA) for two years of the total four years (15.6 and 16.4 for 1998 and 1999 respectively), as well as for the pooled data over four years. The percentage of

variance was observed to be the highest as recorded by SAT II scores during the four years in UCGPA, followed by HSGPA with 16.0% and 15.4% respectively. The rank of SAT I scores was last with a percentage of variance of 13.3% for an equation of single-variable prediction. Nevertheless, in the context of academic success in the years 1997 and 1998, the SAT I score standardised regression co-efficient could not generate a statistically notable predictor.

Shaw et al. (2012) investigated the predictive validity of the SAT through cumulative Grade Point Average (cGPA) in the second year of college based on majors. In addition, they examined the differential prediction of cGPA by college subject across student subcategories (gender, ethnicity, and parental education level). The sample included 39,440 students from 66 colleges and universities in the U.S. The SAT is a better predictor of cGPA in Science, Technology, Engineering and Mathematics (STEM) fields, in multiple correlations of the SAT sections (critical reading, mathematics, and writing) and GPA from 0.61 to 0.63 but computer science, as a part of STEM, exhibited multiple weaker correlations between the majors - 0.50. The undeclared students had the weakest correlations between cGPA and SAT with 0.42. SAT-W among the three SAT sections deemed the most robust predictor in most of the majors. In that context, education was 0.58; however, in the context of biomedical and biological sciences, SAT-M was the strongest with 0.59, and in statistics/physical sciences and mathematics it was 0.59. Apart from that, SAT-CR was evidenced as strongest in the majors of protective and security services, public administration and education - 0.55, 0.55 and 0.53 respectively. The majors for white students were observed to hold the strongest correlations between cGPA and SAT in education and physical sciences/statistics - 0.58 and 0.64.

From a review of the above papers, there is an agreement about the relationship between SAT and GPA. Some studies may have some limitations in terms of their results, however. The limitations generally involve having older data, or that the study itself it is out of date

(e.g. 1990). Moreover, the sample size of Shew et.al (2012) may generate some criticism regarding generalisability in the sense that many of the groups studied had not have enough students to make broad conclusions by subgroup; for example, the Computer and Information Science's group was 48 students. The current study in this thesis, however, attempts to include students from one major and conduct a follow up investigation regarding demography and motivation factors. Lastly, and more importantly, none of the studies addressed in this review explored students who had changed their original subject of study. This could be a very significant factor involved in academic performance at university. The importance, in fact, lies in the fact that there are very few students who fall into this category each year.

3.4.5   GRE

Almost all schools in the United States require the Graduate Record Examination (GRE) from university applicants. The GRE is a common requirement for admission. However, it is only required at the postgraduate level. The main aim of the GRE exam, according to the Educational Testing Service (ETS), is to evaluate verbal reasoning of students, critical thinking, reasoning that is quantitative, and writing that is analytical. Several studies and reports have discussed the predictive validity of the GRE in a variety of contexts. Table 3.11 reviews the findings from example studies:

Table 3-11: Summary of studies that investigated GRE and the GPA.

Study	Place	Sample	Kind of Correlation	Correlation
Kuncel et al 2001	USA	82,659 students	GRE and GGPA	GRE-V = 0.34 GRE-Q = 0.32
			GRE and 1st-year GGPA	GRE-V = 0.34 GRE-Q = 0.38
			UGPA and GGPA	GGPA = 0.30
			UGPA and 1st-year GGPA	GGPA = 0.33
Powers, 2001	USA	1420 students	GRE and 1st-year GPA$_s$	All the three GRE test = 0.53
			UGPA and 1st-year GPA$_s$	The correlation was 0.59
			GRE and UGPA with 1st-year GPA$_s$	The correlation was 0.73
Wang 2013	USA	1083 students	GGPA and GRE	Total GGPA and total GRE =0.28
			GGPA and UGPA	Total GGPA and UGPA = 0.64
Young et al 2014	USA	30 students: 21 part time and 9 full-time	GPA-1st semester and GRE	GRE-Q = 0.39
			GPA-1^{sr} and UGPA	UGPA = 0.13
Wao et al 2015	USA	329 students	GRE and GGPA	GRE-total = 0.112 and 0.39 GRE-Q = 0.114 and 0.46 GRE-V = 0.06 and 0.39

The Table (3.11) indicates that there is a relationship between the GRE and the GGPA, with a difference in the degree of relationship. However, as noted earlier, the GRE is a requirement for studying only at the postgraduate level. The study by Kuncel et al. (2001) involved 82,659 postgraduate students from 1,753 independent samples. This study explored the relationship between Graduate Record Examination results and postgraduates, which named (for graduates in the U.S) different majors. These majors were: humanities, social sciences, mathematics/physical sciences, life sciences and social sciences. They found that the Graduate Record Examination-Verbal (GRE-V) was the greater predictor for Humanities and Life Science majors (0.72, 0.42 respectively). Graduate Record Examination-Quantitative (GRE-Q) scores were the better predictor for social science (0.38) and mathematics and physical science students (0.63). The subject test was found to be a better predictor of first year Graduate Grade Point Average (GGPA) (0.45) and GGPA (0.41). The best correlation between UGPA and GGPA was 0.41 with mathematics and physical

sciences. In contrast, the highest relationship between UGPA and first year GGPA was with life sciences and mathematics/physical sciences, 0.34 for both.

Powers' 2001 article investigated the validity of the GRE for Admission to Colleges of Veterinary Medicine in relation to first year Grade Point Averages (GPAs). The sample was selected from 16 veterinary medical colleges and comprised 1420 students. The majority of the sample was female (70%). In this study, the author found that the greatest relationship was between GPA and the combination of undergraduate Grade Point Averages with GRE, which was 0.73. The correlation between all the GRE sections (verbal (V), quantitative (Q) and analytical (A)) and the GPAs was 0.53 and between UGPA and GPAs in the first year was 0.59. Individually, the validity for the three GRE sections was acceptable. However, the best predictor between the GRE sections for Academic success was GRE-Q then GRE-A (0.49 and 0.46 respectively).

In the same context, Wang (2013) investigated the validity of GRE scores in predicting academic success among postgraduates as measured by Graduate Grade Point Average (GGPA). His study included 1083 postgraduate engineering students in the U.S.A. The researcher identified the predictor variables as GRE, undergraduate Grade Point Average (UGPA), and TOEFL. The criterion variables in this study were Grade Point Averages in the 1st-year (GGPA-1), Grade Point Averages in the 2nd-year (GGPA-2) and total Grade Point Averages (GGPA-tot). In general, Wang found that the best predictor for academic success (GGPA-tot) was UGPA (0.64) followed by GRE-tot with GGPA-tot (0.28). With respect to each section of the GRE test, the strongest relationship with GGPA-tot was found for GRE-Q (0.26).

Young and his colleagues (2014) attempted to find a link between GRE sections Verbal (V), Quantitative (Q) and Analytical writing (AW) and Grade Point Average (GPA) in the first semester for Masters students in business administration. The sample included 30 students

(21 part-time students and 9 full-time students). Their results showed that the greatest predictor for GPA-1st semester was GRE-Q 0.39 and the weakest predictor was GRE-AW (0.07). However, the UGPA had a weak relationship with 1st semester GPA (0.13). A very recent study was completed by Wao et al. (2015). They attempted to investigate the relationship between GRE scores and Grade Point Average for Construction Management students in the United States. They selected 329 students from three universities and Colleges: The University of Florida-Gainesville, Texas A&M College Station and Colorado State University, Fort Collins. They studied the student results from 2009-2014 across every semester and for all the years together. Finally, they revealed a weak relationship between the GRE and Academic success. For all students, the better predictor for academic success was GRE-Q (0.114). The relationship between GRE-V and Graduate Grade Point Average (GGPA) was 0.061. However, the correlation between GRE-total and GGPA was 0.112. For every semester, the highest correlation between GRE-total and GGPA was in summer 2013 (0.39) and it was in the same year and semester that the better relationship between GRE-Q and GGPA (0.49) was found. On other hand, in fall 2013, there was discovered a good connection between GRE-V and Grade points for postgraduates, which was 0.39.

3.4.6   GMAT

The main use of the Graduate Management Admission Test (GMAT) is to help admissions officers of graduate management schools in making admissions decisions. The official website for GMAT, as mentioned earlier, has reported that 4800 management programmes for graduates use their scores in facilitating admissions and more than 300,000 individuals have taken the GMAT test in previous years. Many studies have debated the predictive validity of GRE in varied contexts. Similar to the previous sections, the following Table (3.12) reviews the findings of some of the studies addressing the GMAT.

Table 3-12: Summary of the results of GMAT and university academic success

Study	Place	Sample	Kind of Correlation	Correlation
Sireci and Talento-Miller (2006).	USA	5076 students	GMAT and 1st year GPA	All GMAT = 0.249 GMAT-Q = 0.36 GMAT-V = 0.275 GMAT-AW = 0.038
			UGPA and 1st^t year GPA	UGPA = 0.089
Talento-Miller (2008)	USA	1241 students	GMAT and GPA by Gender	Ma = 0.458, Fe = 0.55
			GMAT and GPA by Major	Buns = 0.425, Engr = 0.467
Kass et al (2012)	USA	72 students	UGPA, GMAT-Q, and GMAT-V with MBA-GPA	0.552
			MBA-GPA and UGPA	0.338

As seen above in table 3.12, very few studies have addressed the relations between GMAT and postgraduate academic performance. They generally confirm a positive relationship. However, the current study will only focus on students at the undergraduate level.

Sireci and Talento-Miller (2006) explored the predictive validity of GMAT in a study that included a sample of 5076 postgraduates from 11 School of Management Admissions in the U.S. They attempted to compare between genders (male-female), race/ethnicity (white, Asian American, African American, Hispanic/Latino and American Indian) with first year Grade Point Average. Their study used zero-order correlations, squared multiple correlations for each analysis, standardised regression weights, and semi-partial correlations for each predictor variable. The results showed a reasonable regression between Graduate Management Admission Test Quantitative (GMAT-Q) and Graduate Management Admission Test Verbal (GMAT-V) and first year Grade Point Average. Those correlations were 0.36 and 0.275 respectively. However, before correction, GMAT-Q, and V each accounted for about 9.3% to 7.5% of the difference in first-year GPA but GMAT-AW accounted for only about 2.1%. However, after correcting GMAT scores, the percentage of variation accounted for more than three times GMAT-Q (36.3%) and GMAT-V (27.5%) but GMAT-AW was still small (3.8%). A review of the means and standard deviations showed

that the women had a lower first year GPA (nearly 1/3 of a standard deviation lower). The white students had the highest score on all variables except GMAT-Q, of which the highest scores were from Asian Americans.

A further study that focused on the predictive validity of GMAT was reported by Talento-Miller (2008). In this study, the data was obtained between 1997 and 2005, and came from 267 studies: 253 studies from schools in the United States and 14 studies were from places in Canada, Europe, and Asia. The number of students included in this study was 1241. The information collected for each student was gender (male-female), native language (English, Western European and other languages), nationality (Western European, Eastern European, Americas and Asia Pacific), major (Engineering and Business) and GMAT scores (GMAT-Q, GMAT-V and GMAT- AW). The results showed that the validity coefficients for GMAT were higher on the non-U.S. programmes (0.516) compared with U.S programmes (0.47). According to variance, the validity of GMAT-Q was large for the European program (0.517) compared with the other scores. Based on percentages, European programmes constitute nearly 25% of GMAT scores. Western European nationalities and languages performed better than other groups (0.27, 0.16 respectively). However, there were no statistically significant findings between groups based on differences in variance.

Another study that investigated the predictive validity of GMAT is that of Kass et al. (2012), which focused on 72 non-U.S students who were admitted onto a Master of Business Administration (MBA) programme at a university in Pennsylvania in the U.S. They used these analyses in their study: sequential regression, correlations and stepwise regressions. The Sequential Regression Analysis was found to be 0.338 between undergraduate Grade Point Average (UGPA) and Master of Business Administration Grade Point Average (MBA-GPA), 0.55 between UGPA and GAMT-G and V, which accounted for nearly 30% of the overall variance in MBA GPA. This result is similar to what Sireci and Talento-Miller (2006) found.

3.4.7   Comments on predictive validity studies and Summary

This section has classified the main measures for predicting future academic success: The International English Language Testing System (IELTS), the Test of English as a Foreign Language (TOEFL), the Scholastic Aptitude Test (SAT), the Graduate Record Examination (GRE), and the Graduate Management Admission (GMAT). This section explained each one of these exams and presented summaries of some studies that examined the predictive validity between these exams and academic success or Grade Point Average. These studies are relevant to the current project that investigates the predictive validity of the current admission standards applied at the College of Education at KFU in Saudi Arabia. In spite of the widespread use of standardised tests at undergraduate and postgraduate level, there remains controversy surrounding their effectiveness as predictors of success, as well as the debate related to how to use them efficiently in the admission process (Berrett, 2013). The Dean of Admissions and Financial Aid at Harvard University headed a committee that examined the test practices and made recommendations on the best use of university admission examinations. The committee concluded that colleges may discuss whether suitable admission decisions can be made effectively without the use of standardised tests. In addition, it stated that standardised admission tests are "incredibly inaccurate" when it comes to measuring academic abilities and how well students are doing at college (Berrett, 2013).

This previous discussion of the literature concerning the predictive validity of pre-university and post-university assessment methods and outcomes at the Higher Education level provided essential background information on past research in pre-university predictive assessment. This may be helpful in terms of explaining some of the results from this study which seeks to examine the predictive validity of the assessment of admission criteria used in the Faculty of Education at King Faisal University. Therefore, in this research, some of the limitations reported by previous studies were addressed through tripartite surveying

tools. Furthermore, some factors affecting predictive strength, such as specialisation or subject change in the first year of university study, as well as gender differences, were taken into consideration. This research thus used a mixed approach in the study of predictive validity to answer the study questions. Additionally, different sources of information such as student scores and the perceptions of lecturers about the current assessment were attained.

3.5   Assessment

3.5.1   Introduction

This section examines assessment in education and the changes introduced to the concept over time. In addition, the section sets out the differences between Assessment for Learning (AfL) and Assessment as Learning (AaL), which is followed by an overview of assessment in Higher Education. Finally, the section will finish by shedding light on the relationship between assessment at the end of High School and assessment at university.

3.5.2   Assessment

In education, the term assessment, or student assessment, has become very popular in academic discussions. The increasing interest has been due to the global view regarding the assessment process and its effect on the performance of students (Brown, 2011). The connection between assessment and education has also been in the spotlight locally in Saudi Arabia, with the effects of interest in educational assessment noticed in the establishment of a Measurement and Evaluation Centre in 2001 (Al-Ghamdi & Gawad, 2015; Alshumrani, 2007; Murshid, 2013). As mentioned in the mission statement on the centre's website, the aim is 'to provide comprehensive and integrated solutions that scientifically measure and evaluate knowledge, skills and aptitude with the purpose of achieving fairness, maintaining quality and satisfying development needs'. www.QIYAS.sa

The concepts of "assessment" and "evaluation" are currently in universal use in educational communities (Taras, 2005; Wiliam, 2010). For instance, in the United Kingdom, the general usage of the word "assessment" denotes evaluation of a pupil's output, whereas "evaluation" denotes assessments concerning subjects or their delivery, as well as the manner of producing such views (Taras, 2005). Over the years, the term "assessment'" has been essentially utilised to illustrate the course of appraising the efficiency of a series of instructional methods as soon as the series is completed (Alaam, 2005; Wiliam, 2011). The measures that govern the learning procedures prior to the completion of the series are not commonly considered as forms of evaluation. French language literature normally regards assessments as facets of the supervision of educational practice, whereas the English language literature considers it to the degree that it was deliberated at all - as merely a feature of effective teaching (Wiliam, 2011). Nonetheless, there has been a growing inclination among English-language research circles to attempt to comprehend activities and methods that are designed to facilitate learning in order to achieve the desired objectives, and those that occur throughout the learning process, as types of assessment (Wiliam, 2011).

It should be pointed out that the terms "assessment" and "evaluation" are used interchangeably in much of the literature, despite the differentiation asserted by particular authors (Alaam, 2005; Pagano, 2007). For example, Davis (1994, as cited in Pagano, 2007) concluded that there currently exists three standpoints. The first categorises evaluation as a subdivision of assessment, while the second places assessment as a subsection of evaluation. The third stance considers both evaluation and assessment as convergent. As for Suskie (2009), he differentiated between evaluation and assessment by asserting that the former uses the latter's information to reach an informed decision concerning whether learners have accomplished the targets set out for them.

While assessment is central in determining the accomplishment of the education system, evaluation is considered an integral part of student learning (Whalley, 2015). Evaluation or

assessment as used interchangeably are terms that refer to an activity seeking to identify student attitudes, knowledge accumulated and skills acquired (Kellaghan & Greaney, 2001). The assessment process is undertaken at the end of instructional acts to determine their impact on the students (Wiliam, 2011). Assessment is the act of coming up with a decision on the quality and degree of learner attainment or accomplishment (Sadler, 2005). There are three types of assessment that are used in education: Assessment of Learning, Assessment for Learning, and Assessment as Learning. These are now discussed in the following sections.

3.5.2.1   Assessment of learning

This kind of assessment is the most used form of student evaluation today. Learning assessment is carried out by conducting tests where students are allocated numbers according to performance (Earl, 2012). The assessment of students takes place at the end of learning units, classes, and a programme, among many other factors that are completed during a specified period (Ibid.). Educators perform summative assessment to gauge the degree of learning that has occurred and to ascertain whether a pupil exhibits all results completely or only partly (Brown, 2011).

On the subject of the underlying principle of summative assessment, Glickman et al. (2009) asserted that it serves as a tool to comply with the educational institution's needs or the district's requirements for teacher responsibility; it also aims to offer a resolution for nonstandard performance, and it presents a basis for removal if needed. Summative assessment is the assessment of learning that is weighed against formative assessment, otherwise regarded as Assessment for Learning. It offers details concerning the outcome's effectiveness or its ability to achieve what it was intended to accomplish.

Glickman et al. (2009) delineated several attributes of summative assessment, including the fact that it involves courses of action that seek to evaluate or grade the degree of learning in a specific timeframe; it makes use of distinct evaluation schemes; and generally aims to be constructive, giving more importance to highlighting what pupils are capable of doing instead of what they are not (Ogar, 2013). An example of the summative assessment in Saudi Arabia includes exams on the various subjects at the end of each term, the High School exam, and the faculty exams for each course at the end of each term or at the end of the academic year.

3.5.2.2   Assessment for learning

Assessment for Learning, or formative assessment, is conducted daily between students and teachers. Formative assessment is a process used by teachers and students during education and feedback to monitor continuing education and to enhance students' achievement in order to achieve the intended learning outcomes (Bennett, 2011)

This type of learning responds to questions posed in dialogues and discussions (Gardner, 2012). Teachers or instructors seek to determine learners' understanding after demonstration and observation. Assessment for Learning, which is also called formative learning, is used to collect views that may help enhance the learning process (Black, Harrison & Lee, 2003, cited in Reinholz, 2015). In formative assessment, students become active participants with their teachers, share learning goals and understand how their learning progresses, as well as the steps they need to take and how to take them (Stull et al., 2011).

The core purpose of formative assessment is to enhance the learning experience of students. There are many studies that demonstrate that effective feedback leads to learning gains. Black and Wiliam (1998) collected more than 250 studies on feedback that had taken place

since 1988, covering all educational sectors. The selection included teachers' assessment, self-assessment, and peer assessment. The results of these studies concluded that feedback produced significant benefits in learning and achievement in all areas of knowledge, skill and types and levels of education.

Darling-Hammond (2008) highlighted the fact that the use of feedback and formative assessment on a continuous basis has positive effects on effective teaching and learning. One of the effects of feedback in formative assessment is to provide students with information about their key strengths and major weaknesses in areas covered by formative assessments. It is also an opportunity to justify the mark or degree assigned, helping students develop self-awareness and motivating them to advance academically (Owen, 2016; Stull et al., 2011).

Sadler (1998), cited three essential conditions that students should be aware of in order to benefit from feedback: (1) the student should know what he is aiming for and what constitutes good performance; (2) the student should know how to relate his current performance to good performance; and (3) the student should know how to bridge the gap between current performance and good performance.

In order to achieve the second and third conditions, Sadler cautioned that students should have assessment skills and that teachers should enhance students' self-assessment skills. Assessment processes can prove fruitful in instances when the learners pose suggestions and the teachers take those suggestions into account, thus enhancing the learning experience (Black et al., 2004, cited in Wiliam, 2011).

In Saudi Arabia, there is still a lack of understanding and application of formative assessment (Alkhalidi & Alturki, 2018; Azizi, 2018). Alkhalidi and Alturki (2018) reported that some female teachers in High Schools in Saudi Arabia doubt the usefulness of formative assessment and feedback. The researchers ascribed this view to a number of factors, including the lack of awareness of the concept of feedback among teachers, while feedback

itself is based on functions that assessment cannot obtain otherwise, such as correction, supervision, and motivation for learners. The researchers also stressed another reason for the teachers' lack of faith in the significance of feedback: the scarcity of Arab studies in this field. According to their questionnaire responses, none of the sampled teachers knew about Arab studies on the impact of feedback on the enhancement of the learning outcomes of the students.

3.5.2.3   Assessment as learning

Assessment can be conducted at times as a learning process. Assessment as learning focuses on students with an emphasis on the evaluation process (Lorna & Katz, 2006). The evaluation process is conducted in a metacognitive way; that is, involving knowledge of one's own thought process (Dann, 2014; Earl, 2012). Assessment as a learning process takes the approach that learning is a two-way process. Ancient learning processes were structured so that knowledge was to be transferred from one knowledgeable person to the students (Lorna & Katz, 2006). Assessment as learning emphasises that learning obtains when a knowledgeable person takes other ideas from the unknowing person (Ibid.). Thus, learning is a process of active cognitive restructuring that happens when ideas are shared (Lam, 2015). In the context of assessment as a learning process, learners are the fundamental bridges between evaluation and acquisition of knowledge. Students are encouraged to relate information to prior knowledge and make sense of such information (Lam, 2015). To successfully perform the aforementioned activities, learners should be critical evaluators of information. The processes of critical assessment and relation of knowledge is a process of metacognition. Through metacognition, learners become experts in monitoring that which they learn. What is achieved from the monitoring process is used to make necessary

adaptations, adjustments, and other necessary changes in the thinking process (Lorna & Katz, 2006).

3.5.3   Measurement:

The streaming of students into Higher Education programmes begins with standardised testing and High School results, which identify student aptitudes and abilities. Higher Education is effective when students improve over time and when they have fair admission to colleges. Without measurement and assessment, it is impossible to know whether students are making any progress or if they will achieve success in Higher Education (Chase, 2009).

Measurement is a pressing issue for all practitioners in the social sciences. Generally, the evaluation of a measure's quality is gauged by how valid and reliable it is. One can refer to the validity of a measure as "the extent to which an operationalisation measures the concept it is supposed to measure" (Bagozzi & Phillips, 1991: 421). On the other hand, De Vellis (2003) defines reliability as the ratio of variance of the attained score that can be attributed to the real score of a latent variable. As such, the purpose of the approach is to adopt tools that can be valid and, at the same time, consistent for all the sampled subjects. Another definition of measurement refers to when objects are attached to numbers or figures, exact and quantitative numbers (Drawnie, 2005, cited in Tshabalala & Ncube (2014). Carmines and Zeller (1979) defined measurement as the 'process of linking abstract concepts to empirical indicants' (p.10).

An attribute of which interpretations vary is referred to as a variable by statisticians, whereas a scale refers to the type of unit by which a variable is measured. One can ascribe the need for measurement scales to the fact that they allow us to change or replace inaccurate words for accurate numbers. There is a limit to what can be done with words but there can be much

less limitation when it comes to figures, which can allow one to carry out specific tasks and operations that cannot be achieved with words. For instance, numbers allow for the utilisation of a range of statistical methods. These figures, in turn, can lead to more precision and impartiality when accounting for behaviour or other issues.

This section has outlined essential information about assessment and evaluation in education. This section also explained the key types of assessment - Assessment of Learning, Assessment for Learning, and Assessment as Learning, and it included a brief description of measurement. The next section presents some background to standardised assessment, which has relevance to assessment in Higher Education.

3.5.4   Standardised assessment

There has been some heated debate with regards to the adoption of standardised testing to evaluate academic accomplishment (Graham & Neu, 2004). In spite of these arguments, standardised testing is adopted in all educational stages from the primary level to college entry tests (Morales & Saenz, 2007). Based on a timeline created by Chappius, Stiggins, Arter and Chappius (2009), the real start of standardised testing was over six decades ago and it has gradually developed and been in use ever since. As stated by these scholars, standardised testing saw its inception in the 1940s to allow colleges to choose their own admissions policies (Ricketts, 2010).

Presently, standardised tests have a number of purposes, such as the gauging of student progress and the assessment of school performance. The use of these tests is initiated by the government, which indicates that they have become a sizable and significant area of interest for public education systems (Ibid.).

Higher Education interchangeably utilises standardised tests in deciding on admissions (Geiser & Atkinson, 2009). The use of standardised testing is founded on two basic suppositions: the first is that this type of test is intended to be as impartial as possible, while the second is that standardised tests precisely evaluate an individual's degree of academic competence (Gawthrop, 2014). Standardised testing primarily alludes to examinations that are homogeneously given and evenly graded (Bond, 1996). They may be standards-referenced tests, wherein an examinee's personal result is weighed against the average distribution of grades in a specific number of individuals, but they may also be criterion-based tests, where an individual's grade is measured against a particular performance scale (Graham & Neu, 2004).

Utilising standardised tests throughout the years has turned out to be more common in various territories worldwide (Alaam, 2005; Al-Ghamdi & Gawad, 2015; Kamens & McNeely, 2010, as cited in Bergmann, 2014).

The discussions and arguments on the subject of standardised tests are ever-present. There is one element of standardised testing that is often contested: the outcomes or otherwise known as the stakes of the examination. As elucidated by Madaus (1988), a test is rendered high-stakes once its outcomes are utilised to make significant decisions about students, teachers, administrators, schools, and/or systems. For instance, according to Au (2007) a high-stakes test could be linked to learners' graduation or, in certain situations, to teachers' or principals' pay.

As reported by Wang, Beckett and Brown (2006), a great deal of evidence seems to confirm that assigning stakes to standardised tests lead to a wide range of unintentional adverse outcomes, including a lack of motivation and engagement, as well as a high proportion of dropout, unscrupulous test preparation, and dubious morals during test management. A frequently ignored, but widely expressed, issue was raised by Linn (2000) in relation to

validity and was focused on whether standardised tests do gauge what they are intended to gauge or whether they actually measure something different, such as the impacts of exam preparation (Bergmann, 2014).

According to Heilig and Darling-Hammond (2008), a number of critics disputing the rationality of standardised testing refer to the pedagogic constraints of such testing and its failure to gauge the spirit of learning. Similarly, several viewpoints are in consensus regarding the significance of the class-time spent on the tests, which amounts to some kind of transaction cost approach to the argument and increases the likelihood that certain instructors 'teach to the test', thus skewing the outcomes. Graham and Neu (2004) posit that these wide-ranging stances do not even begin to capture all the likely viewpoints on standardised testing, but still highlight the multiplicity of opinions in relation to the topic.

3.5.4.1   Advantages and disadvantages of standardised assessment

Drawing on a large body of studies, there seems to be proponents and opponents as regards the issue of standardised testing (Volante, 2006), as noted in the previous section. There are many advantages for the use of standardised assessment in the measurement of learners' achievement. To start with, standardised test scores offer impartial, non-discriminatory, and objective accounts of students' performance, and are directly linked to learning outcomes. The information offered through standardised tests can result in unbiased assessment, as opposed to teacher evaluations which can be influenced by individual decisions on the basis of certain factors, including teacher experience and observation (Alaam, 2005; Phelps, 2008). Another advantage is that standardised tests are measurement instruments because they can be utilised to identify learning issues, help students to move to a higher level, and evaluate students' performance and development in an impartial, standardised manner. While endorsing the standardised test, this should not necessarily lead

to the elimination of imagination and ingenuity in the learning environment or encourage a one-way approach to teaching (Brewi, 2015; Clarke, 2011).

That said, there are some disadvantages associated with the use of standardised assessment when measuring students' achievement. First, this approach to instruction limits and interferes with the syllabus by seeking to adopt the "teaching to the test" methods which take precious time away from non-tested disciplines, especially when high-stakes are associated with final results (Smith & Fey, 2000; Volante, 2006). The other disadvantage relates to the misinterpretation of student achievement levels as a result of dependence on a single assessment method. In fact, standardised tests may not accurately address all student groups, while proficiency levels on tests can be compromised, which could have an effect on how student achievement might be interpreted (Ricketts, 2010). Similarly, the approach weakens education quality by alienating differing opinions, and disabling creativity and intellectual endeavours as a whole (Volante, 2006; Wideen, et al., 1997). Similar to all other approaches, standardised testing has its advantages and drawbacks. However, it is clear from the foregoing that standardised tests do require another supporting tool for the purpose of evaluating some of the higher order thinking skills such as analysis, synthesis and evaluation. This will help maximise the learning potential and improve performance levels of the students, especially in terms of cognition and skills. These criticisms of the standardised assessment would apply also to the Saudi Education system that still uses this form of assessment. The aim of this section has been to give the reader a broad outline of the standardised assessment. The next section explains the alternative assessment types which may be employed instead of standardised assessment, or at least alongside it, in Higher Education.

3.5.5   Alternative assessment in Higher Education

In the literature, a number of challenges have been shown to face conventional assessment, with alternative assessment receiving increasing attention in Higher Education (Chirimbu, 2013; Oliver, 2015). In recent years, there has been a significant rise in the interest in learners' classroom assessment, both in public and Higher Education. The assessment of students' learning outcomes has attracted attention and become an interesting topic in the educational literature because student assessment is an essential component of any curriculum (Quansah, 2018). Therefore, educational reform efforts have developed various assessment methods that can be used to better understand students and to support the achievement of relatively fair assessment standards (Bany Odeh, 2015; Herman, 1997). Numerous unfamiliar terms have been included in educational publications, including real assessment, alternative, and conservative assessments (Elharrar, 2006).

Largely adopted by academics and educationalists in Higher Education is the concept of alternative assessment, which often refers to any kind of assessment of learners responding to a task. One of the major aspects of alternative assessment is the learners' active role in the evaluation of their own performance. Another key feature is how students can use reflective thinking to achieve self-development (Kavaliauskiene, Kaminskiene & Anusiene, 2007).

The coinage of the alternative assessment construct was carried out to describe all those potential tasks and activities which are not formal tests yet might be utilised for judging learning performance, as opposed to the traditional approaches of measuring learners' improvement and performance (Chirimbu, 2013). Thus, alternative assessment represents a shift away from forms of assessment that one might typically see as 'traditional', such as essays and exams.

According to Al-Hajili (2016), alternative assessment has had a significant impact on how educational assessment has come to be known. These radical changes include:

1- Switching from a single testing policy to multi-assessment policies. Therefore, alternative assessment methods may contribute to the fairness of the assessment, as not all students prefer to be assessed using exams. Some prefer to present or to engage in teamwork and be assessed that way. The more diverse the tools and instruments of the assessment, the more positive effects it may have on the final learning outcomes. Students therefore see alternative assessment as "fairer" than traditional assessment methods (Struyven et al., 2005).

2- Shifting from measuring cognitive abilities to measuring multiple skills and abilities. Alternative Assessment may help to assess higher-order thinking skills, such as analysis, structure and critical thinking.

3- Enhancing assessment by making it continuous and integrated rather than being confined to a separate assessment during the middle and at the end of the school year. Post-assessment feedback affects the improvement of learners' learning achievement and reminds them of some mistakes they make in order to avoid them in the future, which renders the alternative assessment a continuous and integrative process (Alkhalidi & Alturki, 2018).

According to several researchers, including McLoughlin and Luca (2001) and Reeves (2000), alternative assessment enables students to be in control of their own learning. It also allows them to gain higher-level thinking abilities and demonstrate their learning potential (Liu, 2007). In terms of the most significant assets of alternative assessment, they entail that assessment should be focused on realistic tasks that demonstrate the student's ability to meet the requirements of communication. Moreover, the teacher and the student should both concentrate on communication, not necessarily on what is right or what is wrong. Another trait is related to the student's ability to contribute to the establishment of the standards in order to effectively achieve the communication activities. Finally, it is important that the

student is offered the opportunity to carry out self-assessment and peer assessment (Ţeican, 2015).

As summed up by Maclellan (2004), the following are some of the benefits gained through the use of alternative assessment in Higher Education. First of all, students are given the opportunity to contribute to the formulation of assessment goals and criteria. Furthermore, students can carry out activities and create an artefact or a product. Another benefit is the possibility to use higher-level thinking and gain problem solving skills. In addition, students are allowed to assess their metacognitive, collaborative and intrapersonal skills, in addition to intellectual products. Other benefits include assessing meaningful instructional tasks, having the opportunity to contextualise in real world situations, and using particular pre-acknowledged standards, which describe criteria for effective performance (Ibid.).

Also confirmed by Huerta-Macias (1995) is the unique characteristic of alternative assessment through its non-intrusiveness as it expands day-to-day curricular-based classroom tasks. The other aspect in relation to alternative assessment is that it allows students to be typically measured on their everyday performance in class. In addition, it offers insights into the strong points and limitations of learners. In addition, it is important that it is appropriately managed because it can be quite a sensitive issue from a multicultural point of view (Brown & Hudson, 1998).

Many studies have, in fact, confirmed how effective alternative assessment can be in terms of developing the skills of learners and enhancing higher levels of thinking, as well as improving mathematical thinking and enabling students to solve problems (Reeves, 2000; Tarawneh, 2004; Oliver, 2015; & Al-Asqah & Dolat, 2017).

In a review by Struyven et al. (2005) that aimed to study students' perceptions of assessment after analysing 36 published studies from 1980 to 2002, the findings showed that some learners consider alternative assessment to be reasonable given its gratifying impact on

students constantly endeavouring to seek knowledge and learn, as opposed to those mostly dependent on a last-minute, crammed effort. Furthermore, as argued by students, alternative assessment is seen as a considerable improvement, which could be due, on the one hand, to the feedback quality students are likely to receive, and to positively communicating staff prospects, on the other. In the eyes of many, openness and clarity were considered major factors of an impartial and effective assessment process. Some issues seemed to arise regarding the dependability of self and peer assessment, despite the students' evaluation of the task (Ibid.).

It is evident from the aforementioned literature that traditional assessment aims to reveal the amount of knowledge gained by students, which is the main objective of using tests as a tool. On the other hand, alternative assessment is not only limited to cognitive aspects of the students, but also aims to develop their various skills and then link them to real life situations. This may enable students to show the ability in problem-solving that is required in reality outside of school (authentic assessment). It is therefore possible to say that the role of the teacher in the alternative assessment has changed from a transferrer of knowledge to a mentor that trains students in how to acquire knowledge (Alaam, 2005).

In modern education, and in many academically developed countries, it is not considered fair to assess Higher Education students solely on the basis of standardised tests. So, if universities want to demonstrate strong exit performance of their graduates and, at the same time, develop dispositions and skills in these graduates which employers demand, the traditional assessment methods may not be appropriate in achieving these goals. If a university shows a limited ability to predict future performance of its students on the basis of their entering educational attainment, to what degree is the way in which the students are assessed a factor in this? Are there issues with the entry assessment mechanisms? Are there issues with the appropriateness of assessment post-entry to a Higher Education institution? In sum, what might be some possible implications for the ways in which learners are assessed

at entry, during and at exit to their university course? For instance, a 'poorly' performing student at the end of High School (on the basis of summative written exams) may flourish and do very well at university because the assessment mechanisms in place (projects, portfolios, oral presentations) permit this same student to demonstrate high attainment.

With this in mind, if the College of Education at King Faisal University intends to prepare students to graduate with skills and knowledge commensurate with the Saudi labour market and in line with the Saudi *Vision 2030*, the traditional methods of assessment based on the examination system may not be sufficient to achieve this goal. As such, the researcher hopes that the current research in part highlights the challenges and disadvantages of the standard tests, as opposed to alternative assessment with its advantages, and should encourage lecturers and professors in the Faculty of Education at King Faisal University, in particular lecturers in the faculty of Education, to adopt various types of alternative assessment tools besides standard testing. This may achieve some sort of balance in the college, especially with regard to the curriculum, assessment methods and the labour market requirements, as well as take into account individual differences among learners, since alternative assessment helps students learn even when they are assessed through feedback.

3.5.5.1   Types of alternative assessment in Higher Education

In general, one can report that there are two types of assessment in Higher Education. The first one is the traditional assessment or standardised assessment, as explained earlier. Traditional assessment refers to exams, including multiple-choice questions, true or false, fill in blanks and essay questions (Dikli, 2003; Liu, 2007). The second type is the alternative assessment, which includes many types such as presentations, portfolios, projects, peer assessment and others (Simonson et al., 2015). The next section sets out some of the types of alternative assessment in more detail.

3.5.5.1.1 Presentation

The presentation has become a common tool in assessment across fields. In Higher Education, the presentation is used as a tool of assessment in many subjects, including education (Al-Ghamdi & Gawad, 2015). Delivering a presentation can either be carried out on an individual level or as part of a group. During the assessment stage, these presentations are generally followed by a written assignment, such as a report or essay. In oral presentations, students are assessed for their ability to provide a summary and show their knowledge regarding a particular subject. As stated by Pearce and Lee (2009), the skills that are typically assessed in oral presentations include understanding of the topic, self-assuredness, succinctness of the answers, quality of the content and answers, ability to react promptly, communication and interpersonal skills, relatedness of theory to reality, effective question management, body language and eye contact, professionalism and clarity of answers. In terms of the disadvantages of presentations, they are usually time-consuming for both student and teacher as the student has to collect information and prepare for the oral assessment, which often involves group work, while the teacher has to spend much time listening to every single student and providing support and guidance throughout the preparation stage (Bhati, 2012). Despite several good features of the presentation method, its use is still limited in Saudi universities as one of the key teaching and learning assessment tools.

3.5.5.1.2 Portfolio

Portfolios are often used in various fields of education for the purpose of tracking the professional development of students (Van Wesel & Prop, 2008). The term "portfolio" in an educational context refers to a group of materials used to keep a record of students' work and achievements during their academic study (Chen, Yu & Chang, 2007; Sandford & Hsu, 2013; van Wesel & Prop, 2008). Portfolios may include a wide range

of materials from student academic activities, including teacher or supervisor reports, rough drafts, graded assignments, photographs, papers, showcase pieces, drawings, reader critiques or summaries, self-reflection pieces, homework assignments, journal entries, test results, peer responses, spreadsheets, and so on (Chen et al., 2007; Sandford & Hsu, 2013). In other words, the portfolio could be likened to a carrier containing a variety of items, works and activities for students to reflect on their academic achievement during their studies. The main purpose of portfolios is to make sense from evidence (Good, 2008) and they are used as an alternative to standardised assessment methods, with many scholars arguing that standardised testing does not appear to provide a clear picture of students' academic achievement (Smith & Fey, 2000; Volante, 2004, 2006).

Portfolios are not limited in their role to providing summary information, but have other features, including offering formative information through evaluation (Barrett, 2006). This means that 'portfolios are not simply another means to an end in determining absolute performance, but are also tools for evaluating the effectiveness of the process and growth of the individual' and academic programme (Sandford & Hsu, 2013, p. 217). Frazier and Paulson (1992) asserted that portfolios can be a source of motivation for students, as well as promoting student self-evaluation and self-understanding. In summarising the advantages of portfolios, they are used to assess academic achievement and to encourage learning, in addition to supporting the efficiency of teaching.

However, some researchers argue that the portfolio instrument is not suitable for oversized classes, not to mention that some students may be reluctant to accommodate this type of assessment (Birgin & Baki, 2007; Chen et al., 2007). Another disadvantage of using portfolios is the amount of time that teachers spend on it to assess a learner's achievement over a period of time, especially in crowded classrooms (Birgin & Baki, 2007; Brookhart, 2008; McDonald, 2012). As a consequence, some researchers have suggested a number of

supportive tools, such as checklists and digital portfolios (Birgin, 2006b; De Fina, 1992, cited in Birgin & Baki, 2007; Lustig, 1996).

3.5.5.1.3   Project

As mentioned in the literature, the use of projects in education can be traced back to 1916 (Metz, 2015). The project, which is one of the alternative assessments in the education field, can be defined in terms of formulating an approach to solving a particular problem, then implementing, observing and reporting the results (Sönmez, 2007, cited in Alacapinar, 2008). Markham et al. (2003, p. 4) have also defined project-based learning as "a systematic teaching method that engages students in learning knowledge and skills through an extended inquiry process structured around complex, authentic questions and carefully designed projects and tasks".

Projects can help students in terms of learning a number of skills, such as engaging in teamwork in the relevant field of study and the ability to share knowledge, skills and feelings (Alacapinar, 2008). There are many benefits for using the project tool in education. Firstly, learners learn how to work as a group to address any potential problems (Alacapinar, 2008; Krajcik et al. 1999, cited in Frank & Barzilai, 2004; Metz, 2015). This may give students the opportunity to share ideas or information and to find a suitable solution for the problem or the best answer to the question. In addition, learners may also learn how to deal with students coming from different backgrounds (Krajcik et al., 1999, cited in Frank & Barzilai, 2004; Railsback, 2002). Similarly, this tool is of benefit for students in terms of dealing with similar problems in real life. Secondly, the project method supports independent learning and students assuming responsibility when undertaking a certain task (Amer, 2015; Metz, 2015). This might help in terms of transforming the educational environment into an attractive setting for learners and boosting students' self-reliance and self-confidence. It should also

be pointed out that projects do generally involve learners in several kinds of tasks which are geared towards meeting their learning needs (Alacapinar, 2008; Krajcik et al. 1999, cited in Frank & Barzilai, 2004), as well as developing higher-order thinking skills, such as analysis and evaluation (Amer, 2015).

However, there are some obstacles that might face students and teachers utilising the project assessment tool in Higher Education. Projects may take longer than expected and require a great deal of preparation and are thus time-consuming for tutors (Railsback, 2002). As learners and educators have little or no experience in projects, they are more likely to opt for an environment that is less demanding or which requires minimal effort on their part (Amer, 2015; Frank & Barzilai, 2004; Railsback, 2002). Some students may face a number of issues when adopting the project method, such as ambiguity, while others may feel frustrated because they have no idea about how and where to start the project and its procedure (Frank & Barzilai, 2004). It could be argued that the problems that face the application of the project in Higher Education lie in the time and effort it takes during the implementation stage and the lack of some students and teachers' skills in relation to the project method. However, these disadvantages can be dealt with through the promotion of training courses and programmes in order to develop teachers in this field through the constant cooperation between the Ministry of Education and Higher Education in the Saudi context.

There are other alternative assessment types that can be used in Higher Education with or instead of the standardised assessment. For example, written assessments, online assessment, written reports, and so on. However, this research referred to some of them as examples of alternative assessment in Higher Education.

3.5.6   Lecturers' perceptions about Assessment:

Interest in Higher Education has increased significantly in recent years. The assessment generally serves various purposes, such as providing information about student achievement, academic progress, quality of teaching and so on. In view of one aspect of the study outlined in this thesis, research into teacher perceptions of assessment is an important issue in educational research. Although it may be one that has wide-ranging implications for policies and practices in Higher Education, few studies have explored faculty teaching staff's perceptions regarding the different aspects of the assessment (Hidri, 2016; Opre, 2015). This can also be the case for their perceptions of predictive validity. The reason may be that interest in studying teachers' perceptions of assessment in education is somewhat recent (Barnes et al., 2015; Brown, 2006).

Studies on assessment concepts provide important insights into how teachers understand the concept of assessment and how these beliefs affect their teaching behaviour, since teachers' beliefs, attitudes, and responses have been shown to influence the quality of the curriculum, teaching practices and the assessment process (Fives & Buehl, 2012). For example, science teachers in Taiwan who view assessment as a major contributor to better learning tend to consider that learning science increases understanding rather than memorisation (Lin et al., 2014). Similarly, Iranian English teachers who believe that assessment seeks to improve learning expressed the fact that they felt less tired and exhausted compared to those who believe that the assessment is irrelevant to the work of teachers and students (Pishghadam et al.,2014). These examples illustrate the importance of the concept of assessment among some teachers in the teaching profession (Brown & Gao, 2015).

Opre (2015) identified four categories of teachers' perspectives of assessment by analysing five studies that addressed teachers' views about the assessment process:

- **The first perspective:** The main objective of assessment is to improve students' levels and the learning process.

- **The second perspective**: Assessment holds students responsible for their learning.

- **The third perspective**: Teachers and educational institutions are to be held accountable for any failure in the level of assessment of the performance of students.

- **The fourth perspective**: This describes the irrelevance of assessment and rejects the idea that the assessment process in education is useful.

Hidri (2016) and Opre (2015) suggested that conducting multiple studies in the field of teacher perceptions of assessment would be useful in terms of analysing the nature and intensity of this relationship between teachers' perceptions of assessment and their teaching behaviour and could make a substantial contribution to the development of this field. In addition, the findings of this type of research could lead to new research trends. By conducting interviews with faculty members at the College of Education at King Faisal University to identify their perceptions of the assessment of public education and Higher Education in Saudi Arabia, the present study may well make one of those contributions that Hidri (2016) and Opre (2015) referred to in a social context, which is completely different from the studies that Aubrey examined in his own research. The present study may contribute some fresh findings about the perceptions of Saudi faculty members regarding the suitability of existing methods of assessment as a predictor of future performance and different aspects of the assessment, and completely clarify other areas, albeit in a different context from the small number of studies conducted previously.

3.5.7   Research on Higher Education assessment in neighbouring countries

Evaluation in the Arab Gulf States may differ from one state to another (Alfassi, 2006). For instance, in the UAE, practical and written exams are still largely considered as the main

tool in terms of evaluating learners in Higher Education (Aljassar, 2012). Paragraph 12 of the admissions systems within the universities of the UAE was dedicated to tests and academic evaluation. Paragraph 12(a) states that all courses should include periodic assessments to gauge the academic performance level, in addition to the final exam. These periodic assessments should be held at a time when students are able to review the grades before the final exam. Additionally, the requirements, dates and deadlines of these periodic assessments should be accounted for in detail in the syllabus of the course (UAEU, 2016).

In Kuwait, however, the evaluation system at the University of Kuwait stipulates that a student should be assessed several times in a single course. According to Paragraph 2 of the evaluation criteria, the final exam score for any academic curriculum ranges between 40 and 60% of the total score, in keeping with the decision of the relevant academic department. The teacher delivering the course should clarify how the marking system works during the initial lecture (KU, 2015).

With regards to the Omani case, at Muscat University, the Higher Education evaluation system provides a continuous assessment of the curricular objectives. The evaluation takes place according to specific standards and not by comparing the student's performance to other peers within the group. For most courses, there are three essential elements contributing to the final evaluation of the curriculum, with none of these elements having to contribute more than 60% to the final evaluation. Furthermore, the last two weeks at the end of each semester are dedicated to the final exam. The final evaluation must cover the entire coursework and the final exam accounts for 40 to 60% of the total overall course mark. The final evaluation may take other forms different from the written exam, including an oral presentation, a seminar or a project, etc., depending on the nature of the coursework (SQU, 2015).

Based on the foregoing information, it can be noted that the majority of universities in the Arab Gulf region assign around 40% to 60% weighting to the final exam. In addition, it seems that there is no clarity in the tools used in evaluating learners, and it usually remains up to the teacher to decide. Arguably, the Omani universities employ a method based on empowering students to carry out comparisons of their own performances as opposed to comparisons with performances of other colleagues, which is the system adopted in the rest of the Gulf universities. Thus, the current research results might be useful for colleges of Education in the Gulf countries that use the same criteria as the College of Education at King Faisal University in the knowledge of the relationship between the admission standards and academic success, since there are exchange of experiences between the faculties of the Gulf of Education.

3.5.8   Research on Higher Education Assessment in KSA

The research in this thesis will be undertaken in one Higher Education institution in Saudi Arabia so it would seem pertinent to review some of the literature on assessment as it relates to Saudi Arabia.

The most notable studies carried out to assess learning outcomes in the Arabic world were summarised by Darandari (2010). The author found that stances on assessment are not clear and comprehensive. Similarly, theories related to student assessment and learning are not incorporated and not well linked at all established and organisational levels. As such, Darandari (2010) put forward a number of recommendations with the purpose of enhancing assessment routines, such as the introduction of well-adjusted, unified, and comprehensive assessment systems and approaches in the Kingdom of Saudi Arabia at the level of classrooms, programmes, and institutions. Even though several workshops have been created by the NCAAA to assist universities and Higher Education institutions with their assessment procedures and practices, the application of new assessment theories has been regarded as a

major challenge for Higher Education in the country. According to the NCAAA approach, this is based on three stages. In the first stage, there is a focus on the promotion of a quality culture in relation to instruction, learning, and assessment. The second stage refers to programme mapping and course planning, with a specific emphasis on assessment plans. As for the third stage, this involves formal implementation of organisational assessment policies.

As stipulated in the regulations of the Ministry of Higher Education in Saudi Arabia, 60% of the end of the course score should be assigned to the final examination, while 40% should be for mid-term examinations, as well as all other assignments and tasks. The majority of assessments in Saudi Arabia were dependent on content examinations that usually use the recall and memorisation technique. As for assessment tools, these are mostly essays, short-answer questions, multiple-choice, and true or false questions. In addition, the curriculum did not include any learning outcomes, while student assessments were not incorporated as a learning experience for the students, which led to the negligence of several vital intellectual, social, and communication skills. In general, little feedback was received by students following assessment, especially at the end of the final exams (Smith & Abouammoh, 2013)

It should be noted that, in the evaluation and examination system issued by the Ministry of Higher Education Arabia, it stipulates the following: (HEC, 2015)

- Article 22: The College Board to which the curriculum belongs specifies a score for the classroom activities to be no less than 30% of the final grade at the end of the academic year.

- Article 23: Classroom assessment score is calculated by means of two methods:

a. An oral or practice test, as well as research or any other types of classroom activities, and one written test as a minimum.

b. A minimum of two written tests.

- Article 24: The College Board to which the curriculum belongs is allowed to include the final assessment in any practical or oral tests, and to specify the grading scores assigned to them from the final test scores.

Articles 33 and Article 34: The teacher should prepare the final test questions and correct them (HEC, 2015). This could also be the task of another person nominated by the College Board upon the request of the department head.

However, the one method that still dominates the methods used in the assessment of student learning outcomes is the examination technique. Most of these examinations vary from essays to true/false questions, multiple-choice, and fill in the blank (Al-Ghamdi & Gawad, 2015).

In Saudi universities, a norm-referenced and summative approach has always been endorsed in student assessment. In fact, assessment is regarded as a procedure to establish some kind of rank ordering among students according to the knowledge, skills and abilities that they have gained once the subject studied or curriculum has come to an end (HEC, 2015; Smith and Abouammoh, 2013). Also important to note is that formal testing has long been the main approach to collect summative information in Saudi Higher Education institutions. Essentially, the focus of any assessment of learning should not only be on classifying students in terms of their performance, but should also involve offering highly useful information about the value and suitability of teaching and learning methods, for instructors and learners alike, as well as the quality of the curriculum designed and its delivery. As such, assessment approaches should be developed concurrently with pedagogical practice and curriculum design and development (Smith & Abouammoh, 2013). The whole process should be part of an inclusive policy intended to enhance the quality of instruction and learning. Almost no evidence of this taking place in Saudi Arabia was actually found, be it

at the system, university, or individual classroom level (Al-Ghamdi & Gawad, 2015; Smith & Abouammoh, 2013).

Nevertheless, according to the new National Qualifications Framework (NQF) in Saudi Arabia, the types of expected student learning outcomes can be categorised into five areas. First is knowledge, which includes the ability to remember, comprehend, and present information, such as knowledge of particular facts, notions, values, theories, and processes. Second are cognitive skills, which refer to the ability to implement conceptual consideration of concepts, values and theories and to implement processes involved in critical thinking and innovative problem solving. The third domain is interpersonal skills and responsibility, which includes the learners' ability to assume responsibility for their own learning and to carry on with their personal and professional development, as well as teaming efficiently in groups and assuming leadership roles when suitable. Learners should also act sensibly when it comes to personal and professional life and behave in the best ethical and consistent manner by observing high moral etiquettes in personal and public settings. As for the fourth domain, it refers to communication, IT, and numerical skills. These relate to the ability to establish effective communication in oral and written form, in addition to the mastery of information and communications technology, and the employment of basic mathematical and statistical methods. Finally, there are psychomotor skills, which involve manual dexterity. This domain is related to a number of specialties, like surgery and performing arts (NCAAA, 2010).

3.5.9   Selected previous studies on predictive validity in Saudi Arabia

There is a paucity of studies addressing the issue of the predictive validity of the admission standard applied in Saudi universities. This is especially true considering that QIYAS (the National Centre of Measurement and Assessment) was launched only recently. The research

in this area is still slow and is very poor, as is clear from the literature reviewed within this study. Only five main studies were found to address the issue of academic success in relation to the scores of High School, Achievement Tests and Aptitude Tests. Those studies are Alshumrani (2007), Alshehri (2011), Alqatei and Alharbi (2012), Murshid, (2013) and Alkushi & Althewini (2020). These are now individually examined in more detail.

3.5.9.1   Alshumrani (2007)

Alshumrani (2007) attempted to identify the credibility of the General Aptitude Test and High School average in the prediction of first year university GPA of the students of the University of Um Al-Qurra in Makah in Saudi Arabia. His study covered 2,170 male students who were admitted to the following colleges: Engineering and Islamic Architecture, Arabic language, Practical Science, Social Studies, Islamic Studies, Islamic Fundamentals and Education. The outcome of the study was that there existed a relationship with a statistical indication between each of the combined secondary average and Aptitude Tests and the first university year GPA, with the interpretation rate reaching 11% of the total variance. The study also found a relationship with a statistical indication between the combined secondary average and Aptitude Test and first university year GPA, with an interpretation rate reaching 11% of the total variance. Secondary average separately explained about 10% of the entire variance of first year university GPA and explained about 11.2% of the entire variance of first year university GPA. The Aptitude Test explained only 1% of the total GPA variance of first university semester and explained about 8% of the entire variance of first university year GPA. However, Alshumrani's research was limited to males at Um Al-Qurra University, as discussed earlier, and revealed two different outcomes and involved some limitations regarding the population targeted, the major covered in the investigation and the research time and context.

3.5.9.2 Alshehri (2011)

Alshehri (2011) conducted his study into male students accepted for study in Al-Taif University in 2009 and who passed the preparation year and spent the first year of their bachelor's degree study in scientific faculties. The study aimed to:

1- Identify the predictive value of acceptance criteria used in Al-Taif University using the success touchstone – the university cumulative average.

2- Rank acceptance criteria (secondary rate, General Aptitude Test and Achievement Test) in accordance with their significance in interpreting the university GPA.

In determining the share of each criterion in the equal ratio used in the trade-off, Alshehri produced the following results:

1- Through the results of correlation coefficients between the GPA and acceptance criteria (High School Percentage, General Aptitude Test and Achievement Test) it was obvious that the High School Percentage and Achievement Test had a medium-level relationship with GPA, 0.48 and 0.40, while the relation was poor between the General Aptitude Test and university GPA, as it reached 0.12.

2- The acceptance criteria used had the ability to predict the GPA of the student, but the entire interpretation rate was not high, being 0.297.

3- The study concluded that ranking acceptance standards according to the interpretation rate of the GPA yielded the following list:

a- General High School Percentage.

b- Achievement Test degree.

c- General Aptitude Test degree.

Nevertheless, Al-Shehri's research was restricted to males in scientific faculties at Al-Taif University.

3.5.9.3 Alqatei and Alharbi (2012)

Alqatei and Alharbi (2012) studied the ability of acceptance standards (general abilities, Achievement Test and secondary average) to predict university performance (for scientific faculties' students) in the first year in the following universities: King Abdul-Aziz, King Saud, King Faisal, King Khalid, King Fahad, Um Al Qurra, Tayba, Islamic University, Imam Mohammed Bin Saud, and Taef University. All of these universities are in Saudi Arabia. The sample of the study was 22,139 students. The study used multiple correlations and adopted Cohen's (1988) standard for explaining the correlation strength. Through analysis of variances in the predictive value, the study found that the variable that contributed best to the prediction of first year GPA in the scientific sections was the Achievement Test followed by the secondary average and then the General Aptitude Test.

The contradicting outcomes between the results of this study and Alshehri's study might be explained by the fact that the Achievement Test is included in the High School performance. Despite the comprehensiveness of this study and the fact that it is very recent and up-to-date, it did not include females, nor did it address the issue of students who changed their major or subject in the first year.

3.5.9.4 Murshid (2013)

Murshid's study (2013) was conducted at the Faculty of Medicine in Tayba University in Al Medina Al Monawarah in Saudi Arabia and examined over 478 students (229 male and 249 female) to determine the ability of acceptance standards to predict the academic achievement

in the faculty. To do so, the researcher used Pearson's correlation coefficient and multiple regression analysis. His results are summarised as follows:

There is a statistically significant relationship between the High School percentage and GPA (0.434). There is a statistically significant relationship between the Achievement Test and GPA (0.213). There is no statistically significant relationship between the Aptitude Test and university GPA (0.039). There is a statistically significant relationship between the High School percentage and Achievement Test (0.243). There is a statistically significant relationship between the High School and Aptitude Test (0.116). There is a statistically significant relationship between the Achievement Test and Aptitude Test (0.420) for all individuals in the sample. However, Murshid's population was only limited to students in the Faculty of Medicine and included male and female applicants.

3.5.9.5   Alkushi & Althewini (2020)

In a study by Alkushi & Althewini (2020), they aimed at identifying the predictive truth of admission criteria in King Saud bin Abdulaziz University for Health Sciences. The study was applied to 1,595 students. As for the data, they were analysed with polynomial logistic regression and multivariate linear regression. The findings demonstrated that all admission criteria were predicting college assignments but only 21.1% of variance. The best predictor was the general aptitude test. This study differs from the current study in that it was applied to the health field, as well as targeting only the first-year criterion for learners.

These studies reveal different outcomes and limitations regarding the targeted population, the major covered in the investigation, and the research time and context. The variety of outcomes and limitations strongly suggest the need for further investigation regarding which of the three admission items greatly contribute to students' academic success at undergraduate level. Further, based on the literature being reviewed, no study was found to

address the predictive validity among students who changed their major after the first year. In addition, QIYAS has only included females since 2010 (eight years after applying it to males). A potential gap is that none of the studies reviewed above covered female students in Education within their population. Therefore, investigating the predictive validity of academic success that includes males and females would be of great interest to researchers and policy makers within the Arabic region.

With regards to the predictive validity which targets specific majors, none of the studies being reviewed had covered the discipline of education in any of the Saudi universities. This strengthens the need for investigating the predictive validity of the admission standards at a School of Education to explore which of the criteria for admission is the best predictor for academic success.

Being an instructor and member of staff at the Educational College at King Faisal University (KFU) the current researcher of this study became more interested in the admission standards at the Educational College, since many students tend to change their major during their university study. Further, from his experience teaching courses within the Education degree, some students with high scores in High School and QIYAS tests often fail some of the modules within the programme. This gives further importance to investigating what contributes substantially to students' success at the Educational College at KFU and what, if any, impact this might have on the ways in which undergraduate education students are assessed.

3.6   Summary

This chapter aims to give the reader a broad summary of the literature related to the predictive validity of the admission standards that are used in Higher Education and also assessment in Higher Education. It covered validity and the prediction of future academic

success i.e. SAT, IELTS, GRE, GMAT and TOEFL. The chapter also discussed assessment and standardised assessment in Higher Education. It has covered some important alternative assessments. Then, this chapter presented research on Higher Education assessments in Saudi Arabia and the neighbouring countries to Saudi Arabia. Finally, the chapter concludes with the previous studies on predictive validity in the Kingdom of Saudi Arabia.

Chapter 4   Research Methodology

4.1   Introduction

The aim of this thesis is to investigate the predictive validity of the admission standards used in the Education College in King Faisal University in Saudi Arabia. This qualitative and quantitative (mixed methods) study intends to answer the following key questions.

1- Is there a statistically significant relationship between the student accumulative assessment score at the end of High School and the accumulative rate at the end year 2014 in the College of Education in King Faisal University at Ahsaa at Saudi Arabia?

2- Is there a statistically significant relationship between the student mark in the General Aptitude Test and the accumulative rate in the College of Education in King Faisal University?

3- Is there a statistically significant relationship between the student marks in the Achievement Test and the accumulative rate in the College of Education in King Faisal University?

4- What is the degree of the following independent variables' ability in terms of interpreting the accumulative rate of each department within the College of Education in King Faisal University?

- The accumulative rate in the High School.

- The marks of the General Aptitude Tests.

- The marks of the Achievement Test.

5- Which admission standards have the greatest ability in terms of predicting success in the College of Education in King Faisal University?

6- Is there a statistically significant relationship between the score of the General Aptitude Test, the Achievement Test and the High School percentage with respect to the following variables: gender and change of major?

7- What are the Education College Lecturers' perceptions of the current admission standards used at the Education College at King Faisal University (The accumulative rate in High School, the marks of the General Aptitude Tests, and the marks of the Achievement Test)?

8- What opinions do Education College Lecturers have on the relationship between assessment undertaken at the end of High School and assessment within the college of education?

The main purpose of this chapter is to outline the research methods that have been used and to illustrate the procedures employed to collect and analyse data. This chapter also discusses the literature underlying the selected methods and the reasons for the procedures for data collection. This chapter will be separated into two sections. The first section presents an overview of the literature on the research methodology (qualitative and quantitative approaches) and the data collection methods that are used with each individual type. The second section of this chapter will focus on the research setting and participants, tools for data collection, data analysis issues of validity and reliability, and matters of ethics.

4.2 Research paradigms

In any research, the adopted paradigm can be seen as the major vehicle shaping how such a piece of research can be carried out (Cooper, n.d.). While undertaking research, there are three elements that constitute the foundation for research paradigms. According to Guba (1990) and Denzin and Lincoln (2003), these are the ontological, epistemological and methodological approaches. The definition of the research paradigm is the description of the

research model upon which the ontological, epistemological and methodological dimensions are grounded. While a framework is related to ontology, epistemology raises queries about the instrument used to gain insights and the certitude and trust emanating from those insights. As for methodology, it is the means through which the ontology and epistemology can be studied (Denzin and Lincoln, 2003). In addition, ontological research is focused on the issue of individual consciousness. It also looks into whether reality is external to the person or is the result of personal experience. By providing an input into the establishment of the paradigm, the epistemological dimension is largely interested in identifying if knowledge can be gained or if further experimentation is required.

According to Porta and Keating (2008), in directing researchers to the methodologies, the paradigms can be categorised into different ways to demonstrate belief systems. As stated by Bassey (1999), two potential paradigms can be identified for categorising the intricacy of concepts. These are the positivist and the interpretative models. For Tashakkori and Teddlie (1998: 7), positivism is considered a paradigm that "bases knowledge solely on observable facts and rejects speculation". Similarly, Pring (2000: 36) expands by asserting that the experimental design is a feature of the scientific paradigm that educational research might seek to imitate. In short, according to the positivist line of thought, all genuine knowledge is centred on sensory experience, which can largely be promoted by observing and experimenting (Ahmed, 2008; Cohen et al., 2007). This is the central theme of knowledge in the positivism approach and it is relevant to the current research, which is intended to investigate the predictive validity of the admission standards in the College of Education at KFU in Saudi Arabia and in which the researcher used a database that showed the real result (observable facts) of the students' achievements.

Unlike with positivism, proponents of the interpretivist approach claim that the social world can only be assumed from the perspective of those participating in the current issue being examined (Cohen et al., 2007). There seems to be heavy reliance within the role of

interpretivists on the researcher's participation within the study subjects. By this, meaning is assumed both through interaction among respondents and between the researcher and the members of the study community, while ascertaining the research findings takes place through interaction with the study subjects (Given, 2008). From an interpretivist point of view, by making informed assessments of people's perceptions and opinions regarding a research issue, researchers can gain further insights. In fact, this is the central theme of knowledge in the interpretivist approach and is relevant to the current research, which is designed to identify some Lecturers' and professors' perspectives in the College of Education at KFU in Saudi Arabia in terms of the QIYAS tests and their predictive validity.

4.3   Research methodology

A paradigm assists as the vehicle for the research and it outlines the research to be undertaken (Cooper, n.d.). According to Tashakkori and Teddlie (1998: 7), positivism relies mostly on knowledge drawing on recognisable facts and discards conjecture about root causes and the meanings that people make from their lived experiences. There are several procedures adopted in the process of data collection (tests, interviews, questionnaires, observations, etc.) for the various research methodologies (Zohrabi, 2013). In a research methodology, a number of tools are utilised in given domains of research activity (Nachmias and Nachmias, 1996). Even though there is no true or incorrect methodology, the academic or researcher should select the appropriate method available given the research questions being asked (Ibid.). It is possible to categorise data collected as "qualitative" if they are presented in word form and refer to certain conditions, people, or events related to a particular situation or issue. On the other hand, they are classified as "quantitative" when they come in the form of figures, statistics or measurements that seek to offer a precise account of observations (Braun & Clarke, 2014; Bryman, 1988, 2015). As such, one can infer that the most widely used

categorisation is between the quantitative and qualitative methodologies (Smeyers, 2001). As claimed by Denzin and Lincoln (1994) and Allwood (2012), it is possible to use both qualitative and quantitative methodologies appropriately within any research philosophy, which is the intent in this thesis.

According to Boyd et al. (1985), selecting the research method is dependent on the nature of the researchable issue. The authors also posit that when adopting a particular scientific method, the researcher has to be fully aware of a number of challenges, including the level of his/her involvement in the use of the findings, inaccurate measurement tools, the likely effect of the measurement process on the findings, time constraints in achieving the findings, issues encountered when using experiments to assess hypotheses, and the intricacy of the problem investigated. The current researcher endeavoured to take these challenges into account in the conducting of the present research.

Porta and Keating (2008) assert that there is a clear difference in the literature between the two methodologies, even though they are not mutually exclusive, and are both applied in many studies. According to Bryman (2015), the mixed methods research approach is utilised as basic shorthand for an approach to research where quantitative and qualitative research methods are combined within same the study. Bryman found that the mixed method approach had increased threefold in less than ten years from 1994 to 2003 (Bryman, 2008). Zohrabi (2013) agrees with Bryman by arguing that mixed methods have recently increased in prominence. In the present research, the mixed method approach was adopted.

4.4   The setting for the study

The current research was undertaken in the Education College at King Faisal University (KFU) in Saudi Arabia. This setting is accessible to the researcher because he is a lecturer

at the same college. The Education College, which is located in Ahsaa City in the east of Saudi Arabia, is one of the oldest colleges in the kingdom.

4.4.1   **Research Sampling**

According to Cohen et al. (2008), one can decide on the quality of a piece of research not only by the suitability and relevance of the methodology and instruments used, but also by the appropriateness of the sampling approach. However, a question that could be challenging in this respect is the size of the sample needed for the researcher to meet the requirements of their research study (Bouma and Atkinson, 1995; Bryman, 2015). As stated by Cohen et al. (2008), there appears to be no definite answer on what constitute a good sample size, given that the appropriate sample size is completely reliant on the aim of the study and the population type (p. 100).

In general, there are two major types of sampling in the education field - namely the probability sample and the non-probability sample. According to the probability sampling method, each member of the whole population has a chance in terms of being included in the sample (Cohen et al., 2007; Given, 2008). The purpose of this type of sampling is to achieve representativeness, which refers to the level at which the sample truly reflects the whole population under study (Teddlie & Yu, 2007). According to Berg and Lune (2012), the use of probability sampling techniques is largely present in quantitatively based studies. Similarly, there will be less risk of bias associated with the use of a probability sample than a non-probability sample. The common basic types of probability sampling include random sampling, stratified sampling, cluster sampling, and multi-phase samples, or sampling using multiple probabilities (Teddlie & Yu, 2007).

Non-probability sampling is a prevalent method in qualitative research and in this method, researchers depend on their judgment in their sample selection (Creswell, 2007; Given, 2008). According to Given (2008), participants chosen by means of the non-probability sampling method are selected because they fulfil pre-arranged requirements. Examples of the more prevalent kinds of non-probability sampling methods include convenience sampling, snowball sampling, and purposive sampling (Berg and Lune, 2012; Bouma and Atkinson, 1995; Cohen, Manion & Morrison, 2007; Given, 2008). Overall, in the case of qualitative sampling cases, they are usually small since scale or estimates of statistical significance are not required. The aim of a qualitative study is usually to gain deep and comprehensive insights. Similarly, there could be problems in terms of the management and analysis of a massive amount of in-depth data (Algarfi, 2010). As stated by Ritchie and Lewis (2003), the small-scale approach can only work if the researcher has a robust sampling approach.

The sample for this study was selected because it includes students of the Faculty of Education from whom the researcher aims to identify the predictive value of admission criteria. The choice of the period 2010-2014 is underpinned by the fact that the tests of the Measurement and Evaluation Centre used to be restricted to male students from when it was first established in 2002 through until 2009. In 2010, the measurement centre tests in Saudi Arabia started to involve female learners for the first time. As such, one aspect of this study is intended to shed light on gender differences in scholastic achievement. The sample was selected from the 2010 school year, i.e., from the year that female students were first included. In addition, the researcher chose to research a four-year period for his study (from 2010 to 2014) in order to examine a complete set of data for the period of study required for full bachelor's degree students, which may give a clearer picture of the study sample. The faculty members were selected from the Faculty of Education at King Faisal University to take part in a series of one-off individual semi-structured interviews because they were the

individuals most familiar with the admission criteria used in the Faculty of Education. Apart from having a good background in the achievement of students in the Faculty of Education, these teaching members of staff were also easy to approach and recruit because the researcher himself is a member of the Faculty of Education at King Faisal University.

As a result, it was quite straightforward to communicate with and reach out to other colleagues. The available members were selected at the time of the interview and the interviewer sought their approval and informed consent to carry out the interview. Eight faculty members, all male, were interviewed, but no female teachers were because female educators are very difficult to approach for cultural reasons in the Kingdom of Saudi Arabia. The distribution of faculty members according to their academic functions was as follows: two Lecturers, three Teaching Assistants, two Associate Professors and one Professor. The below table (4-1) shows the distribution of members according to their academic positions.

Table 4-1: The academic rank of the participants

Participant	Academic rank
P1	Assistant Professor
P2	Associate Professor
P3	Assistant Professor
P4	Professor
P5	Lecturer
P6	Associate Professor
P7	Lecturer
P8	Lecturer

The academic rank of the participants is identified here. However, the researcher has not provided the specific job title because of the risk of disclosing the identity of participants.

As such, the researcher is not in a position to give the full academic title of interviewees because it may then be possible to infer the name of the participant.

The above represents purposive sampling that focuses on deliberately choosing particular cases that will offer as much information as possible in relation to the questions being investigated (Pagano, 2007). In this type of non-probability sampling, the aim is not to make generalisations about the wider community, but to enable a detailed explanation of systematically chosen datasets and sources. (Algarfi, 2010; Given, 2008).

4.5   Data collection

According to Cameron (2011), mixed methods research, which has witnessed an unprecedented increase in prevalence and popularity during the last decade, is frequently referred to as the third methodological movement. Studies have increasingly become dependent on the use of both qualitative and quantitative data that are simultaneously gathered, analysed and then discussed. One commonly held belief is that the adoption of various kinds of methods to gather data and the acquisition of that data using different participants, such as learners, teachers, programme staff members, etc., can contribute to the enhanced validity and reliability of the information collected and its explanation (Zohrabi, 2013). Questionnaires, interviews, classroom observations, focus groups and tests can be stated as major instruments used in mixed method studies (Tashakkori & Teddlie, 2003; Zohrabi, 2013). It can be safely said that these multiple methods of collecting data can complement each other and, as such, maximise the validity and trustworthiness of the information conveyed (Tashakkori & Teddlie, 2003). As stated by Zohrabi (2013), generally speaking, quantitative data are acquired by means of closed-ended questionnaires and tests, while qualitative data are largely obtained using open-ended questionnaires, interviews, focus groups and classroom observations. This research used mixed methods for many reasons. Using a variety of methodological tools in data collection may lead to greater

validity. Additionally, this allows scholars to answer their questions from different perspectives. This gave the researcher more depth to investigate the predictive validity of the current admission standards in the Education College of King Faisal University.

4.5.1   Interviews

One of the most common tools in qualitative studies is interviews (VanderStoep & Johnston, 2009). According to Kvale (1996: 14), an interview can be defined as 'an inter-change of views between two persons conversing about a theme of mutual interest'. As confirmed by Cohen et al. (2007), there are three purposes that interviews may serve. First of all, it is possible to use them as the principal method of collecting data that could have a direct impact on the research aims. Secondly, it is possible to utilise them to test hypotheses or to yield fresh ones, or as a clarifying tool to help find variables and links, and this is indeed the reason for using the interview tool in this research. Finally, it is possible to use interviews in combination with other tools.

The main types of interviews used in education research include structured (or standardised), unstructured, and semi-structured. A structured interview can be identified based on the questions, which are usually closed-ended, and the number of questions and the wording of the questions, which are the same for all participants (Frankfort-Nachmias and Nachmias, 2007). This kind of interview is intended to ensure that every interviewee is provided with the exact same number of questions that must also follow a similar structure (Algarfi, 2010). It has been argued that this form of interview offers more objectivity and is easier when it comes to analysis, though it might not offer flexibility (Frankfort-Nachmias & Nachmias, 2007). Useful though it may be in certain contexts, the researchers in this kind of interview do not generally have the opportunity to follow up on interviewees' responses, and, thus, such a method does not allow for much insight into aspects related to the topic under study

(Alharbi, 2011). As a consequence, researchers are more likely to miss out on valuable details (Gall, Borg & Gall, 1996).

The second kind of interview in education research is unstructured. Unlike with structured interviews, in an unstructured interview format, the researcher has a plethora of themes or subjects to discuss with the interviewee. However, in this type of interview, which is used by some qualitative researchers, the process is largely participant-led (Braun & Clarke, 2014). As for the drawbacks of this type of interview, it requires more time and effort and is harder to analyse. In addition, the validity may be called into question as a result of the subjective nature of the research (Alharbi, 2011). Despite such criticism, this type of interview offers flexibility and could be utilised to gain further insights into the issues under study (Kidder et al., 1986).

The third type is the semi-structured interview, with interviewees being given a limited number of open-ended questions, but where, most of the time, participants' answers are probed by interviewers, who would encourage them to give more detail and explication where it seems fruitful to do so. It should be pointed out that these data are usually dealt with using the qualitative approach (Harris & Brown, 2010). In the semi-structured interview, there are open-ended and closed-ended questions. This indicates that not all questions are pre-designed or formulated prior to the interview meeting (Algarfi, 2010). One can safely say that semi-structured interviews can be used as a flexible approach to enable the researcher to raise new questions or issues during the interview depending on what the participant says. It thus gives the interviewer the chance to gain further insights (Ibid.). In addition, this kind of interview offers standard information across participants but with a more detailed perspective, as opposed to the structured interview (Robson, 2000). In this research, the researcher used a semi-structured interview with open-ended-questions in order to probe respondents' answers and gain further insights into their perceptions about the admission standards in the College of Education at King Faisal University. Besides, it is

important to examine at some length the opinions of the College of Education Lecturers about the perceived relationship between assessments undertaken at the end of High School and those within the College of Education. A review of the literature reveals that this type of interview was used in many other studies to investigate both predictive validity, measured quantitatively, as well as lecturers' and professors' perceptions (Billows, 2007; Elharrar, 2006; Nolan, 2011). The researcher interviewed 8 academics who are currently teaching a number of education courses to different academic levels at KFU. Such academic members included Lecturers, Professors, Assistant Professors and the Dean of the College of Education.

As for the interviewing procedures at the Faculty of Education, the interviews were conducted in the researcher's office at the Faculty of Education at King Faisal University. It is a large office with an air conditioner, coffee maker, water cooler, three comfortable chairs, and two tables. One interviewee asked to have the interview in his own office and the interview was conducted in this manner at his request. At the beginning of each interview, the objective of the research and its importance were clarified to each individual interviewee. All respondents were asked for permission to record the interview and everyone agreed. It was also clarified that each interviewee has the right to ask for clarification if a question is unclear or requires further explanation. The researcher stated the following to each respondent: "You are entitled to request that you do not complete the interview for any reason without any inconvenience or repercussions as a result".

The researcher met faculty members at the Faculty of Education at King Faisal University, and many of these members were colleagues whom they have known for more than ten years. All interviewees are Saudi nationals who understand Saudi culture and context in general and specifically the systems and structure within its education system, in particular at the higher education stage. The researcher was somewhat apprehensive about not getting useful and in-depth research information. This apprehension was manifest in two aspects; the first

relates to the answers to some questions using some common terminology by some colleagues by virtue of our common knowledge. This is known as the silent assumption, (Mazzei,2007) which refers to their belief that the researcher know the answer, so they do not delve further into the interview questions, and they are satisfied with saying (as you know). This happened to more than one subject whom the researcher was asking to cite examples or clarify the answer so that they could continue to speak and explain their answers in the way they think about them, and not assuming what the researcher knew. One of the positives in the process was that the researcher managed to gain some responses from my subjects to my questions after requesting clarifications. They also cited examples of some of their general terms. In those moments, the researcherwas acting as a researcher who wanted to know all the details related to the research, and not an academic colleague who who finds a brief clarification sufficient.

The second aspect was the meeting of the Dean of the college; it was an interview that was cautious on my part due to the position and status of the Dean and it lasted 35 minutes due to the Dean's busy academic diary. It was also important that the researcher kept to the agreed allocated time when the appointment was made to interview the Dean. It was sufficient and covered all the interview questions and the researcher was careful not to ask anything that was disrespectful or challenging to the role of the Dean. It should be noted here that the interview with the Dean of the College had to be booked by the Dean's Secretary. After that, the reason for the meeting and the expected duration of this meeting were made clear. Then, it was a waiting game until approval from the Dean is gained. Once approved, a sample of the questions that will be posed during the meeting with the Dean needs to be handed to his secretary to be checked before the meeting. During the meeting, the researcher has to adhere to the question format that has previously been submitted to the Dean's Secretary and the duration agreed beforehand. The order of the interviews was as follows: to start with, the Dean of the college, given his academic background and status. The meeting

was aimed to explore the latest developments in the assessment in the university in general and in the college in particular. Perhaps offer any differences in the approach to the interview given these participants were of a different academic rank (Professor, Assistant Professor, and lecturer). Regarding the data obtained by the participants, no one was able to access the interview data except for the researcher and the research supervisor who had access to the data presented in the results while writing the study. The data obtained from this study were not passed to any third parties. There was a secret number on each record of the interview, which only the researcher was able to open.

The interviews took place during the academic year 2016-2017 and the average interview time was between 40-50 minutes per meeting. The eight interviews, which took place within the space of a month, were then transcribed in Arabic, which took another month. In the end, to address respondent verification, each interview was returned to the interviewee to check the truthfulness of the information recorded and that it was representative and accurate as regards their viewpoints. Each interviewee was then invited to make any modifications that they deemed appropriate, but, ultimately, no interview transcript was modified by any interviewee. The qualitative sample had to be of sufficient size to include all those who would be able to speak about issues of assessment and predictive validity. The individuals interviewed hold a range of roles within the Education department, represent a range of academic ranks, and have potentially different years of experience in Higher Education. Furthermore, participants may or may not themselves have taken the QIYAS tests at the end of High school. The study results may not generalise to other Education departments in Saudi Arabia because this study was specifically conducted in one Education department, that of KFU. Moreover, the people who were interviewed were those that were available to volunteer and the researcher cannot generalise these results to other faculty members.

4.5.1.1   Interview questions:

In terms of topic, the interview questions in this study were divided in a number of ways: firstly, there was a link to the relationship between the interview questions and the research questions.

The main purpose of the interview questions was to answer the research questions - particularly the last two questions of the eight research questions within this thesis, which were:

- What are the College of Education lecturers' perceptions about the current admission standards used at the College of Education at King Faisal University in Saudi Arabia (The accumulative rate in High School, the marks from the General Aptitude Tests and the marks from the Achievement Test)?
- What opinions do the College of Education lecturers have about the relationship between assessment undertaken at the end of High School and the assessment within the College of Education?

A number of factors influenced the development of the interview questions. Firstly, some questions emerged from the researcher's own experience of Higher Education in Saudi Arabia and his knowledge of the specific setting that formed the context of this thesis. Secondly, questions were also informed by the outcomes of the quantitative analysis part of the thesis. Thirdly, the interview questions were also closely informed by the review of the literature. Specifically, interview questions were designed by reviewing previous research (e.g. Hill, Storch & Lynch, 1999; Kerstjens & Nery, 2000) that addressed the predictive validity of some admission criteria in Higher Education, and which used the interview in the research design. The review also included research that touched on the perceptions of teachers or lecturers about evaluation in general education or Higher Education (e.g. Brown & Gao, 2015; Harris & Brown, 2010; Pishghadam et al,. 2014). In addition, conversations

took place between the researcher and his supervisor regarding the relationship between the interview questions and the research questions. Peer debriefing with the thesis supervisor, who played devil's advocate and challenged the suggested questions, was also undertaken. The first draft of interview questions was then discussed with the researcher's supervisor to gain further feedback and suggestions that helped with the identification of the final interview questions. With the help of such resources, the researcher was able to select and formulate interview questions appropriate to the current research.

The interview questions were initially developed in English and were discussed and agreed with the researcher's supervisor to obtain feedback and suggestions. After that, interview questions were translated into Arabic and then the Arabic translations of the interview questions were submitted to two Saudi Arabian PhD students studying Applied Linguistics at the University of Southampton who validated the accuracy of the researcher's translations.

The outcomes from the pilot study yielded three suggestions. Firstly, participant one suggested using the term 'evaluation' instead of 'assessment' because it was, in their view, the most relevant terminology in the Arabic language. Secondly, participant two proposed deleting question two, which was: 'What modules in the different year groups do you teach?' He said, 'this is because the question is not linked to the objectives of the research'. Finally, participant 1 suggested amending the wording of the last question (nineteen) from: 'Can you tell me anything else about the use of assessment in High School and in College?' to: 'Do you want to add anything about the use of the assessment in High School and in College?'.

4.5.2   Institutional Data:

Institutional data is the information created, collected, transferred, or registered by the university or for conducting university work, such as information recorded on paper, or in

electronic, audio and video formats, for example. These institutional data are entered over extended time periods and may have been entered by different people. However, such data is and was not collected by the researcher nor entered into the system. One of the challenges faced by the researcher in this respect is that there is a possibility that there might be an error when entering the data because it has not been validated. This is an internal database and there are no external checks carried out on it. This indicates that the internal reliability of the database depends on the lecturers and faculty members. In the UK, external examiners have the possibility to check and modify the marks of a sample of students, but this system does not yet exist in Saudi Arabia. We can say that one limitation of the use of institutional data is that, in some cases, it may not be externally verified. It should be noted that this is not an issue of trust, and this is merely a practical consideration of the internal database.

Furthermore, one of the other challenges of institutional data is that there is a possibility that some data entries may be incomplete. Therefore, the current research is limited to accessing data for those students who have completed the four years, which means that the current data in this research does not include any student who temporarily suspended their studies or left the university before the end of the 2014 school year for any reason.

Another challenge facing institutional data is that of Internet attacks. Therefore, one of the IT management functions within institutions of Higher Education is to guide institutions in their quest for data security, information systems and networks. Therefore, the institution must provide assurance of information security as part of its important activities (Grajek, 2013).

In his study which included 19 Higher Education institutions in the U.S., Yanosky (2009) mentioned that about 30% of the institutions studied were subjected to at least one attack via the Internet in the two years preceding the study and that the incidents were more common in postgraduate institutions. Therefore, as soon as the current researcher received the

students' data on a flash drive from the Office of Admission and Registration at King Faisal University, the researcher printed it on A4 paper and placed it in a securely locked briefcase so that if the flash became infected by any viruses, there would be a written copy stored safely in password protected format for the purpose of the research.

With all these challenges, one of the benefits of such data is to show trends over time when changes can lead to specific strategies to improve those data, as well as to determine what some university needs may be based on what these data show. Institutional data can help the institution look for progress and quality for its academic business. As Yanosky (2009) points out, effective institutional data management relies on the use of information to encourage student success, reduce student dropout rates, and support transparency, but attention must first be paid to the quality and validity of these data.

At King Faisal University, students take their tests. Each lecturer corrects the tests for the course he is teaching and sets grades for each student based on the evaluation criteria used at King Faisal University (See Appendix E). A faculty member at the Faculty of Education reviews how the grades are allocated. The faculty member (teacher of the course) monitors the scores electronically through the system. A copy of each student's record shall be kept at the Deanship of Admission and Registration at King Faisal University.

4.5.3   The Database

The other method involving quantitative data collection procedures refers to gaining access to anonymous students' files from the Admission and Registration Office. These data included High School percentages, the General Aptitude Test scores, Achievement Test scores, and cumulative last year college Grade Point Averages, and these were extracted from the students' files, for both male and female students who were admitted in 2010 and

graduated in 2014 from the College of Education at KFU. Two major drawbacks of studies that employ databases refer to time and financing (Grady, Cummings & Hulley, 2013). With regard to the procedures for obtaining the sample for the current research, it should be pointed out that, after the researcher had obtained all necessary approvals to conduct this research from the University of Southampton and King Faisal University, the researcher requested permission to gain access to the data of students (male and female) from the registration office at King Faisal University. The researcher identified himself as a Faculty member at the Faculty of Education at King Faisal University and explained to the official that he was currently studying for a PhD at the University of Southampton in the UK. The researcher clarified the importance and objectives of the current research to the administrative member of staff at the Office of Admission and Registration. It was also confirmed that the researcher would use the data for current scientific research only. The data were obtained within three working days of the researcher's request, which was in April 2017. The data included the following: the grades of the students in High school, the grades for the Achievement Test, the general capacity test scores and the cumulative averages of the students at the end of the bachelor's degree in 2014 for both male and female students. The data did not include students who left the Education College at KFU before the end of the academic year 2014 nor any students who had not graduated by the end of the academic year 2014. In addition, the database retrieval did not reveal any personal information about individuals (names, ages, etc.). It just yielded information about assessment scores. No one was able to access the data except for the researcher and the research supervisor who had access to the data presented in the findings during the writing-up of the study. The data obtained from this study were not passed to any third party. The total sample accounted for 693 students, 309 male, and 384 females.

4.6   Data Analyses

To answer the current research questions, data was collected through interviews (Research questions 7 & 8) and from one admission database (Research questions 1-6). Multiple regression analysis and correlation coefficients between the independent variables (the accumulative rate in the High School, the marks of the General Aptitude Tests and the marks of the Achievement Test) were used to answer Research Questions one to six. Two multiple regression models, one for each criterion variable (the accumulative rate in the College of Education in King Faisal University), were developed. This procedure allows us to determine which one of the predictors contributes (the accumulative rate in the High School, the marks of the General Aptitude Tests and the marks of the Achievement Test) the most to the criterion variable (the accumulative rate in the College of Education in King Faisal University). This may help to ascertain the most valuable predictor in predicting College academic success. From the review of the literature on predictive validity studies, many authors have used these statistical tools (e.g. Al-Malki, 2014; Alshumrani, 2007; Kaiser, 1982; Kerstjens & Nery, 2000). However, research questions seven and eight were answered using interviews. As was explained previously, the researcher used semi-structured interviews with open-ended questions. The researcher interviewed eight academic members of staff who taught a number of education courses at different academic levels (year of study). The sample included: Lecturers, Professors, Assistant Professors, and the Dean of the College of Education, who were all interviewed separately and, prior to the interviews, were introduced to the overall objectives of the interviews. The strict confidentiality of the interviews was confirmed by the researcher, who also stressed the anonymity guarantee to the interviewees when reporting and that pseudonyms would be used to protect the identity of all participants and settings. The researcher sought formal permission to have all interviews taped because of the benefits of such a process. According to Gall et al. (1996), tape recorders can be a very useful tool whose benefits outnumber those of notetaking when

collecting interview data for any kind of research. More specifically, recording minimises the inclination of interviewers to go for an unconscious choice of data that corresponds with their own preconceptions (Alharbi, 2011). Through an audio recording, it is possible to provide a complete verbal account, which can later be studied in a thorough manner and could prove more effective than data in the form of written notes (Alameri, 2017). It should finally be noted that recording interviews may speed up the interview process and, therefore, does not cause interviewees to experience boredom.

The key method in this research is to undertake a data analysis that includes the extraction of substantial relevant meaning from the various interviews. According to Glesne and Peshkin (1992), qualitative analysis refers to the process of understanding what has been heard, witnessed and read by the researcher throughout the data collection process. The researcher is thus required to make hypotheses and categorisations, as well as search for patterns that will be useful for data interpretation and theory development, and in order to relate data to each other and to other data. An argument was put forward by Creswell (2007) that researchers have to undertake several preparatory steps in order to analyse a qualitative project. The initial steps include organising and preparing the data, followed by a minimisation of the data to a codified structure. The next step is to arrange the codes into various themes according to the ideas that need to be developed or investigated in the research. A final procedure is to classify data with rich themes in several sub-themes or by means of a discussion, figures or tables. According to a number of researchers, including Creswell (2007) and Cohen et al. (2003), there exist a host of techniques in the examination of qualitative data, including thematic analysis, grounded theory, content analysis and constant comparison, among others. Thematic analysis refers to the process of recognising, examining, and producing patterns (themes) within data, which involves the organisation and description of the researcher's data set in exhaustive detail. Nevertheless, Boyatzis (1998) claims that it may even involve the interpretation of a number of features of the

research topic. In addition, thematic analysis enables the researcher to provide an exact account of the associations between concepts and to match them to the replicated data (Marks and Yardley, 2004). The adoption of thematic analysis in this research could be useful in terms of linking the different concepts and attitudes of the academics, and comparing these with the database, developing connections that will be formed in various situations and at various periods throughout the lifetime of the project. There are unlimited possibilities for interpretation in the research, though the researcher acted on the proposals of Braun and Clarke (2006) by employing their six steps to thematic analysis, which were undertaken in this thesis. In the next paragraphs, each individual step is explained from the point of view of the literature, followed by an explanation of the procedures which the researcher followed.

First, the researcher should familiarise themselves with the data. It should be pointed out that constant repetition of reading the data is seen as beneficial to fully understand the context. It may be worth transcribing verbal data such as interviews into written form so that a thematic analysis can be carried out (Braun and Clarke, 2006). According to Riessman (1993), transcription is extremely helpful to start familiarising oneself with the data. In addition, as argued by some researchers, transcription should be looked at as an important stage when analysing data in the context of the interpretative qualitative paradigm (Bird, 2005).

In the current research, the researcher applied the same steps as put forward by Braun and Clarke (2006). The researcher used manual coding within the interview transcripts. The first step of the interview analysis involved reading the interview transcriptions in full. These were read several times to gain an overall picture and understanding of the contents and ideas within. This procedure was carried out for the Arabic version as well as the English version of the text for each participant.

The second step refers to the generation of initial codes. Once the researcher has been able to grasp the content of their data, produced a preliminary list of ideas included in the data, and established the items most relevant to the research, then the first codes from the data can be produced. According to Boyatzis (1998: 63), codes can be defined as "the most basic segment, or element, of the raw data or information that can be assessed in a meaningful way regarding the phenomenon". Part of the data analysis is the coding process (Miles & Huberman, 1994), which is divided into two types, namely open coding and axial coding. Open coding refers to the allocation of a code label, a linguistic item or expression found in the transcription or text (Strauss & Corbin 1998). Axial coding refers to the creation of themes or classifications by combining codes or label sets for words and expressions (Richards, 2009). In the current study, the researcher undertook line by line reading looking for key words or phrases inductively which were underlined and assigned a code which ultimately formed a list of codes that emerged from the reading. The researcher used this coding procedure and generated some 56 data codes. Within the initial stage of qualitative analysis, an independent coding exercise was undertaken with the researcher's supervisor on a sample of pages from one transcript. Both the researcher and the supervisor independently coded and then compared their codes and the supervisor undertook peer debriefing in an attempt to challenge the researcher's interpretation of the transcript in order to strengthen the trustworthiness of the data.

The next step, which involves searching for themes, re-focuses the analysis at the wider level of themes, instead of codes. It aims at arranging the various codes into potential themes, and then organising all the required coded data excerpts within the recognised themes (Braun & Clarke, 2006). At the end of this step, candidate themes are collected and so are the sub-themes, as are all relevant data extracts that have been coded (Boeije, 2002). In this third step, the researcher tried to collate the codes that he interpreted as being related to one another. Then, the researcher refined the codes by eliminating the less frequent ones and

those that were seen to be unconnected to the current study. Finally, the researcher sought to generate a theme for all codes that were linked or related in some manner in terms of their properties and features due to their assigned code/label.

The fourth step is focused on reviewing themes. This stage, which starts once the researcher has produced a group of candidate themes, involves the refinement of such themes (Braun & Clarke, 2006). It is important that the researcher reads all the collected excerpts for each theme before considering whether they seem to produce a logical form (Alhojailan, 2012). At the end, the researcher should have a fairly good idea of what their various themes are and how they fit together.

In the fourth step, the researcher reviewed the themes and discussed them with his supervisor and justified how he had arrived at the themes. The researcher then re-drafted some of these themes. For example, the subject of QIYAS tests as admission criteria was modified to insufficient admission standards.

In the next step, which includes defining and naming themes, the researcher should be able to clearly decide the relevant themes for the research (Miles & Huberman, 1994). According to Braun and Clarke (2006), names have to be short and need to provide the reader with an explanation of what the theme refers to.

In this step, the researcher identified the final 'titles' of the themes. The seven identified themes are as follows: (1) Traditional methods continue, (2) Lack of variety of assessment, (3) The relationship between High School assessment and Education College assessment already exists, (4) Knowledge is priority, (5) Insufficient admission standards, (6) Changing subject, (7) and Girls' performance. The researcher reviewed these themes with the supervisor in their final form. The researcher tried to make the titles as short and specific as possible and attempted to explain what each theme referred to.

The final step is to produce the report. This phase brings the process to an end, with the researcher embarking on the finishing touches, including a report of the interview analysis that would eventually lead to the final analysis and writing of the report (Braun & Clarke, 2006). In this step, the researcher wrote up the results based on the themes found and set out the range of ways in which participants referred to the individual themes. Indicative quotes are used to illustrate each theme.

4.7   Triangulation

Triangulation can be utilised by researchers to validate their findings and give them more depth (Brannen, 1995). According to Foss and Ellefsen (2002) and Lockyer (2004), the term 'triangulation' can be defined as the dependence on more than one research approach when examining a specific question. Triangulation has the ability to offer a multidimensional perspective as it signifies the employment of various research approaches. In addition, Ritchie and Lewis (2003) showed that, with triangulation, it is expected that the adoption of diverse sources of information has a major role to play in terms of confirming and enhancing the clearness or accuracy of a research finding. As such, one kind of data (often quantitative) is utilised to confirm another type of data (usually qualitative) (Brannen, 2004). Furthermore, Yin (2003) enumerated the advantages of triangulation, including the variety of sources of evidence and instruments of analysis. Triangulation also enables the researcher to deal with a plethora of historical and behavioural topics or themes. Similarly, it causes the case study to become more conclusive and precise. Triangulation was used in this study in order to strengthen the validity of evaluation data and findings. Triangulation was undertaken across tools for data collection, interviews and from one admission database, and also across interviewees from various educational disciplines, as well as job backgrounds.

4.8   Member checking

Member checking is another approach used to improve internal validity in qualitative research, and it refers to presenting data and unclear answers back to the participants from whom they were taken and checking with them if the findings can be considered credible (Merriam, 2016). A major research function of member checking, which is also called informant feedback or respondent validation, is to help enhance accuracy, authority and reliability. It aims to evaluate intentionality and rectify factual errors, as well as provide participants with the chance to contribute further information or to put information on record. In short, it pertains to ensuring the appropriateness and acceptability of the studied matter (Cohen, Manion & Morrison, 2007). In addition, the process allows the interviewed individuals to appraise the collected material for accuracy before the research enters the analysis stage, which will ensure that the statements are true (Zimbicki, 2007). According to Lincoln and Guba (1985), while triangulation can be used in the process of checking data, member checking, and components of reliability, are useful as a check on respondents' production of information. In this thesis, participants who are interviewed had the opportunity to check the accuracy of interview transcripts.

In the current research, the interview questions were discussed in detail with the researcher's supervisor to gain valuable feedback and take account of the research questions and relevant literature. The Arabic translations of the interview questions were submitted to two PhD students studying in the language field at the University of Southampton, and with both Arabic and English language proficiency (they are originally from Saudi Arabia). This procedure is followed in order to validate the accuracy of the interview questions' translation. The researcher also pilot interviewed two PhD students at Southampton University who work in the education field in Saudi Arabia and explained to them the purpose of this study. Afterwards, the researcher asked them to comment on the suitability

of the interview questions, given their knowledge on assessment within the education system in Saudi Arabian universities.

4.9   Validity and Reliability

The notions of validity and reliability are significant concerns that should be taken into account by educational researchers. Validity is defined by Garett (1937) as the fidelity with which a test is intended to measure (quoted in Rumsey, 2013). As for Bachman (2000), validity is described as a foundation upon which the consideration of values, their uses and effects are based. As such, validity is a requisite for both qualitative and quantitative research approaches. It is possible to address validity in qualitative research by adopting an honest, profound, rich and diverse data collection process when engaging with participants in the research and through the triangulation of data collection instruments (Cohen, Manion and Morrison, 2007). According to Cohen et al. (2007), enhanced validity can be achieved in quantitative data by means of meticulous sampling, suitable instrumentation and adequate statistical handling of the information. In other words, the researcher is not attempting to engage in research offering full validity as much as seeking research that has a higher level of validity. In any quantitative research, there is a measure of standard error that is embedded yet has to be recognised. Contributing to a degree of bias in qualitative data are a number of factors, including the partiality of respondents, their attitudes, their beliefs and their viewpoints. It is important, thus, to consider validity as a matter of degree and not as a complete state (Gronlund, 1990). In so doing, researchers aim for the minimisation of invalidity and the maximisation of validity.

One of the core fundamentals in the educational research process is the reliability of the data and findings. According to Nunan (1999), reliability generally deals with how consistent, trustworthy and replicable the findings achieved through a research study are. As such,

reliability is focused on the issue of how likely it is, or indeed whether at all, the findings of a piece of research can be repeated (Bryman, 2015). In contrast, achieving similar results in quantitative research is rather forthright because information or data is presented in a statistical format. In qualitative approaches to research, however, it would be somewhat challenging and difficult to obtain similar findings, as the data would normally be of narrative nature and largely subjective. As such, Lincoln and Guba (1985) suggested that rather than gaining similar findings, one should focus more on data reliability and consistency. In so doing, researchers should not aim to achieve the same results, but rather come to the understanding that, on the basis of the data collection processes, the findings and results are consistent and reliable. As stated by Merriam (1998), it is possible for the human instrument to become more dependable once given training and hands-on practice. Nevertheless, one can be consistent in qualitative research by trying to assess the quality of the research, which is dependent on the aim of creating understanding. In spite of the argument about the significance of dependability in educational research, according to Moss (1994) and Golafshani (2003), it seems that validity and reliability go hand in hand, and that the notion of reliability can be sufficiently satisfied through the concept of validity, which means that reliability is a result of validity in any educational research. As confirmed by Chioncel et al. (2003) and Cohen et al. (2008), validity and reliability in qualitative research can be achieved by reducing the level of bias to a minimum, for which triangulating via several methods and addressing trustworthiness can be the best answer. As far as this research in Saudi Arabia is concerned, mixed methods (qualitative, quantitative), interviews, and the database are all triangulated as a means to uphold the validity and reliability of the findings. Furthermore, the researcher sought expert and informed opinion in the quantitative field at the University of Southampton on the research questions and methodology, as well as discussing the statistical software programmes to be used in this research. Finally, the researcher outlined the participants involved in the interview procedures, the duration of the

interview, the number of respondents, and the tool that was used to store the interview data. In the eyes of the researcher, this is to clarify the research procedures and is therefore indicative of consistency. The researcher also analysed the qualitative data and listed the tools used to answer all the research questions. During the analysis of the results, he mentioned the studies consistent or conflicting with the results. This refers to validity and may have significantly contributed to ensuring a high level of research credibility.

In addition, as a lecturer in the College of Education I am essentially an 'insider' and I am aware of my position. I declare my researcher subjectivity as a person that has an interest in predictive validity within Education. Based on my experience between 2002-2012 as a lecturer at the College of Education at KFU, there are differences in student performance on the basis of entry level. I have worked at the College at KFU for many years and worked with a number of such students. I have concerns about students who come to the university from High School with high scores but do not do as well as we would wish via their GPA. I have been in conversation with my colleagues about the performance of students and what we know about them when they join the university. In my mind I have some ideas about how we might address performance that is less than expected and improve that prediction. My task during the qualitative analyses is to make sure I stay open to the opinions that participants share in relation to admission and assessment and not reveal any biases I may have. I ensured what I included in the transcription is exact through member checking, that is, by allowing individuals to verify the accuracy of the transcripts and, if they desire, amend or remove any of the content. Furthermore, in the process of early analysis, I went through peer debriefing, which has been defined as the 'process of exposing oneself to a [disinterested] peer to explore aspects of the inquiry that otherwise might remain only implicit in the inquirer's mind' (Cooper et al., 1997: 2). This technique helped the researcher to have a tool for establishing the credibility of a project. Regarding Cooper (1997), there are four main aims of using peer debriefing. Firstly, peer debriefing helps keep the questioner

valid by investigating for meaning, bias, and understanding. Secondly, it provides an opportunity to examine research hypotheses that may be emerging in the questioner's mind. Thirdly, it gives an opportunity to enhance and test next steps in the emerging methodological design. Fourthly, it allows researchers a space for catharsis. During this thesis, I went through peer debriefing with my supervisor, who 'played devil's advocate' and challenged my interpretation of the data.

The qualitative sample used in this research is a purposive sample which focuses on deliberately choosing particular cases that offer as much information as possible in relation to the questions being investigated (Pagano, 2007). In this kind of non-probability sampling, the aim is not to make generalisations about the wider community, but to enable a detailed explanation of a systematically chosen dataset and sources. (Algarfi, 2010; Given, 2008). Thus, this data may be of interest inside the KFU College of Education. Furthermore, the findings might apply to and transfer to other Education Colleges in Saudi Arabia, and, more generally, be of interest to Education Colleges who are interested in the predictive validity of admission criteria.

4.10   Ethical concerns

Ethical consideration pertains to how to treat those contributing to research and how the data is dealt with after having been collected from the participants (VanderStoep & Johnston, 2009). In highlighting the significance of ethical considerations in any research effort, McDonough and McDonough (1997: 54 as cited in Rumsey, 2013) confirmed that there are a number of benefits to ethics, including the protection of research validity through the achievement of good data by acknowledging that data taken from respondents is theirs and its use can only be allowed with their consent, and also the protection of the research informants by rules of privacy, unanimity and permission in terms of specific uses of the data.

To ensure the validity of the research study, the researcher will adhere to the following strategies: firstly, he will have to follow the ethical framework as set out by the University of Southampton. The researcher received the approval from the ERGO committee at the University of Southampton (see Appendix A). Secondly, the College of Education at the University of Southampton sent a request through the researcher to the Saudi Culture Office in London in order to obtain approval from the researcher's sponsor to carry out the research. Thirdly, after receiving approval from the Saudi Culture Office, including the approval from the researcher's sponsor, the researcher obtained approval from the Education College in KFU to allow the researcher to access the database for research purposes only and to carry out the interviews. After receiving all the necessary approvals, the researcher made sure that the database would not be used to reveal the names of the students. The database was only used to check the relevant data needed for the research regarding those students. Therefore, the database will not reveal any information about individual students or any other party, as the researcher is only interested in gaining access to information about assessment scores. In addition, written permission to record and transcribe interviews was obtained from all the lecturers and professors who have taken part in this study. Informed consent was sought and received from all academics who were interviewed. They were assured that their participation is voluntary and that they were free to withdraw from the study at any time without penalty.

4.11   Summary

This chapter has characterised the research methodology for the present mixed methods research. Some examples of qualitative and quantitative research were presented, as well as details of data collection approaches, triangulation and member checking methods and their use in this research. In addition, this chapter has described the setting of the present thesis

and the procedural tools used for data analysis, and, furthermore, discussed validity, reliability, and ethical concerns.

Chapter 5   Pilot study

5.1   Introduction

A pilot study is one of the important stages in a research project. This chapter aims to describe the significance of and steps involved in executing a pilot study by using an interview method. Van Teijlingen and Hundley (2002, p. 33) define a pilot study as 'a small investigation to test the feasibility of procedures and to gather information prior to a larger study'. A pilot study can be used in qualitative, quantitative and mixed methods projects (Gumbo, 2014). Pilot studies are a mini picture of a full study and involve the specific pre-testing of a particular research instrument, for example, questionnaires and interview schedules (Baker, 1994). A pilot interview is conducted because it increases the researcher's awareness of the interviewees' background and predilections, for example, their preferred means of communication and their willingness to participate in academic studies (Mikecz, 2012). A pilot study provides a chance to check the clarity and validity of the interview questions and the comprehension of questions related to the research topic. It can also help the researcher become familiar with the procedures in the interview protocol. Feedback from pilot study interviews will then be reflected in the thesis methodology.

In the current research, there are many reasons to use a pilot study before proceeding to the final project. It is difficult to obtain rich information that is useful to the research topic without interview practice (Mikecz, 2012). Interview practice in advance with individuals familiar with the topic and setting is essential to make the most of this method. This may require the researcher to prepare the interview questions and make sure all questions exhibit clarity and relevance to the main topic. In addition, a pilot study is beneficial and valuable for the researcher in order to have experience of how to manage and prepare an interview with participants such as professors and a Dean.

5.2   Current pilot study

In the current research, the pilot study has been adopted in order to identify any potential practical issues from using the interview technique with the intended participants. For example, some of the participants may give a brief answer that does not fully answer the question, which leads the researcher to request more information, such as: 'Can you tell us more about it?' or 'Please, give as much detail as you can' in order to encourage the participant to expand further. In addition, it was important to have insight into the participants' views on the translation of the interview questions from English into Arabic and the accuracy of such translation, as well as the correlation of interview questions to the research topic and whether there were any questions that needed to be modified or deleted. It was also important for the researcher to gain experience of the interview techniques. Finally, in this pilot study, the researcher strove to avoid any waste of time and effort. As mentioned by Blaxter et al., 2010, a pilot study is a crucial aspect of academic research in order to prevent loss of time, effort and money.

5.2.1   Pilot Study Preparation

First, after completing Chapter Four, which is the methodology chapter, and after reviewing a number of studies in the literature associated with the current study in different countries, the researcher formulated 19 questions to be used in the interview process (See Appendix B). The researcher then translated these questions from English into Arabic. The researcher also consulted his supervisor for feedback about the interview questions, and suggestions. The researcher prepared separate pilot interviews with two postgraduate students specialising in linguistics attending the University of Southampton and who are teaching members of staff at a College of Education in Saudi Arabia. These individuals are familiar with the procedure for assessment in both High Schools in Saudi Arabia and within an

education department in Higher Education. The researcher explained the research topic, objectives, research questions and hypotheses to the participants. In order to check the clarity and internal validity of the interview questions, the researcher went through the interview questions and received comments on the appropriateness of interview questions for the purpose of the study. It should be pointed out that pilot participants' suggestions were useful and helped to improve the interview protocol. The major modifications proposed by the participants included:

- Participant 1 suggested using the term 'evaluation' instead of 'assessment' because it is the most relevant terminology in the Arabic language. Since the study applied to Arabic speaking participants, it would be more appropriate to use the correct term. Otherwise, inaccurate answers might be collected.

- Participant 2 proposed deleting question (2), which was: 'What modules in the different year groups do you teach?' This was because the participant said 'this is because the question is not linked to the objectives of the research'.

- Participant 1 suggested amending the wording of the last question (19) from: 'Can you tell me anything else about the use of assessment in High School and in College?' to: 'Do you want to add anything about the use of assessment in High School and in college?'

Finally, the researcher asked the participants to review the translation from English into Arabic, which was considered correct, except for 'assessment', which was changed into 'evaluation' (see Appendix C). It should also be noted that the researcher had to prepare for some generally accepted practices that related to mobile phones and individuals entering an office, specifically that, in Saudi Arabia, mobile phones ring regularly and I had to accept that if somebody takes a phone call or if somebody comes into the interview room during my interview, then I need to stop the interview, pause the audiotape and resume after the interviewee has finished his call. The interviews took place in the interviewees' offices or in

the researcher's office in the Education College at King Faisal University and this depended on what the participants preferred. The researcher sat down across the desk from each participant, which, again, is a generally accepted procedure when in the office of another.

Chapter 6   The Findings

6.1   Introduction

The aim of this chapter is to present the results following data collection. The findings for research questions 1-6 were drawn from the existing database as mentioned in the methodology chapter and data were collected from 693 students enrolled between the years of 2010 and 2014 on a four-year bachelor's degree in Education. Quantitative data are presented by research questions which are as follows:

1.   Is there a statistically significant relationship between the student accumulative assessment score at the end of High School and the accumulative rate in the College of Education at the end of year 2014 at King Faisal University in Ahsaa, Saudi Arabia?

2.   Is there a statistically significant relationship between the student score in the General Aptitude test at the end of High School and the accumulative rate at the College of Education at the end of year 2014 at King Faisal University?

3.   Is there a statistically significant relationship between the student score in the Achievement Test at the end of High School and the accumulative rate at the College of Education at King Faisal University?

4.   What is the degree of the following independent variables' ability to interpret the accumulative rate of each department within the College of Education at King Faisal University?

-    The accumulative rate in the High School.

-    The marks of the General Aptitude tests.

-    The marks of the Achievement Test.

5. Which admission standards have the greatest ability to predict success at the College of Education at King Faisal University?

6. Is there a statistically significant relationship between the score of the General Aptitude test, Achievement Test, and High School percentage with respect to the variables of gender and change of major?

6.2 Descriptive Statistics

The descriptive statistics were generated using SPSS software Version 25. The data below describe the sample in terms of gender, changes in subject major, and their scores on the High School assessment. The data being described were sampled from a total of 693 students from the 2010-2014 cohort.

Table 6-1: Gender distribution of Sample

	Frequency	Percent
Male	309	44.6
Female	384	55.4
Total	693	100

A total of 693 students took part in the study and the majority (55.4%) were female. Only 52 (7.5%) of the total students changed their majors after admission, and all of these individuals were female.

Table 6-2:Distribution of students depending on major subject

	Frequency	Percent
No change major	641	92.5
Changed Major	52	7.5
Total	693	100

6.3   Quality of students based on admission standards

Overall, the students in the sample had very high scores in their cumulative High School assessment with an average score of 91.2 (SD = 7.1), but with female students having higher mean scores compared to their male counterparts (female mean = 96%, SD = 4; male mean = 85.1%, SD = 5.2). Similarly, students had a higher General Aptitude test score with an overall mean of 75.4% (SD = 7.7) but with female students having higher General Aptitude test scores compared to male students (female mean = 77.8%, SD = 7.5%; male mean =72.3%, SD = 6.7). Furthermore, students performed in line with the average in their Achievement Tests at the end of High School (mean = 79.9%, SD = 8.0). It is important to note that the Achievement Test score relates only to female learners. On the other hand, students had an average GPA of 3.5 (SD = 0.9), with female students having higher mean GPA compared to male students (female mean GPA = 4.0, SD = 0.6; male mean GPA = 2.98, SD = 0.96) see tables 6-3 and 6-4 below for details.

Table 6-3: Students' scores on the High School assessment, General Aptitude Test and achievement.

	N	Minimum	Maximum	Mean	Std. Deviation
High School degree	693	68.50	100.0	91.2	7.1
General Aptitude Test.	693	55.00	100.0	75.4	7.7
Achievement Test	384	50.00	99.0	79.9	8.0
The accumulative rate at the College of Education at King Faisal University in Alahsa (GPA)	693	1.000	5.0	3.5	0.9

Table 6-4: Gender differences in the High School cumulative assessment score, General Aptitude Test, and GPA

	Male					**Female**				
	N	Mean	Min	Max	SD	N	Mean	Min	Max	SD
High School degree	309	85.15	68.50	97.59	5.25	384	96.00	75.77	100.00	4.00
General Aptitude Test	309	72.33	56.00	94.00	6.73	384	77.81	55.00	100.00	7.49
GPA	309	2.98	1.00	4.71	0.96	384	4.00	1.74	4.95	0.60

Table 6-5: Changing major subjects, High School cumulative assessment score, General Aptitude Test score and GPA

| | **Change Major** | | | | | **Not Change Major** | | | |
	N	Mean	Min	Max	SD	N	Mean	Min	Max	SD
High School degree	52	94.07	75.77	99.94	6.27	541	90.93	68.50	100.00	7.10
General Aptitude Test	52	71.22	55.00	87.00	7.24	541	75.70	55.00	100.00	7.60
GPA	52	4.00	1.74	4.76	0.69	541	3.50	1.00	4.95	0.94
Achievement Test	52	69.9	50.0	84.0	6.9	332	81.4	55.0	99.0	7.0

From table 6-5, it is clear that students who changed their majors after admission had higher High School cumulative assessment scores and GPA (High School cum. Mean = 94.1%, SD = 6.3; GPA = 4.0, SD = 0.69) compared to those that did not change their majors (High School cum. Mean = 90.9, SD = 7.1; GPA = 3.5, SD = 0.94). On the other hand, students who did not change their majors on admission to the university had higher Achievement Tests scores and General Aptitude Test scores (Achievement Test mean = 81.4%, SD = 7.0; General Aptitude Test score = 75.7%, SD = 7.6) compared to those that changed their major subjects on admission (Achievement Test mean = 69.9%, SD = 6.9; General Aptitude Test score = 71.2%, SD = 7.2).

6.4 Statistical analyses to answer research questions

The statistical analyses to be undertaken in this subsection are mainly dependent on the nature of the research questions asked and the nature of variables with regards to whether they are continuous or otherwise. Research questions 1-3 are concerned with the bivariate relationships between end of High School cumulative assessment scores (SGPA) and the university cumulative assessment score (UGPA); the General Aptitude Test mark (GAT) and UGPA; and the Achievement Test Score (ACT) and UGPA, respectively. On the other

hand, research question four (RQ4) focuses on the joint effects of the three independent variables (SGPA, GAT and ACT) on GPA. For research questions 1-3, bivariate correlation analysis technique is believed to be more suitable given that the questions require measuring the strength of the association between two quantitative, continuous variables (see Field, 2013). To estimate the joint effects of the three independent variables (Question 4), a linear multiple ordinary least squares regression analysis is the most suitable analytical technique. More precisely, the researcher used a linear hierarchical OLS model. This is against the backdrop that multiple linear regression models are suitable for estimating relationships between a continuous dependent variable and many independent variables.

Given that the Pearson's correlation and linear regression analysis techniques are all linear models, they assume particular conditions for accurate results. Broadly, Field (2013) summarises them as follows:

 (i) Linearity assumption: This assumption requires that the outcome variable (i.e. UGPA) is linearly related to each of the predictors (i.e. SGPA, GAT and ACT).

(ii) Normal distribution assumption: This assumption requires that residuals are normally distributed.

 (iii). Homogeneity of variance or constant variance of residuals: This assumption requires that variance in the outcome variable is equal across the different values of the predictor variable.

These assumptions are tested before the correlation and regression analyses are undertaken to test whether the data conforms to the assumptions.

The test of assumptions revealed a mixed scenario. Foremost, the scatter plots for linearity of relationship show that the relationship between SGPA and UGPA is linear while the relationships between GAT, ACT and UGPA are indicative of no clear relationship pattern.

(see Appendices R, S, and T). With regards to normal distribution of residuals, the data is slightly negatively skewed but does not deviate very much from a normally distributed dataset (Appendix U). The P-P plot also shows that there is a very small deviation from the norm (Appendix V). Therefore, evidence from Appendices U and V shows that the data meets the normal distribution assumption. In terms of homogeneity of residuals, Appendix W does not show any trend in the residual distribution and, hence, it is concluded that the residuals are independent of each other. In conclusion, the data conforms to most of the key assumptions, including normality and constant variance, but two relationships violate the linearity assumption.

Different transformations were tried for the reported two non-linear relationships. The transformations tried included quadratic, log, cubic and square root but not all yielded a good fit in comparison to the linear models. Moreover, some of the transformations, if applied, would make interpretation of the results difficult to understand. It was decided to maintain the original variables without transformation. This implies that non-parametric correlation (Spearman's) is estimated for research questions 2 and 3, while Pearson's correlation is estimated for research question 1. On the other hand, to ensure valid results in the prevailing circumstances where two relationships violate the assumption of linearity, the bootstrapping technique is invoked to generate robust confidence intervals and significance tests of the model parameters (Field, 2013). Further, standardised Beta coefficients are used to answer research question 5 (RQ5) that requires the identification of the most influential independent variable(s) in predicting UGPA. Finally, an independent samples t-test is used to answer research question 6 (RQ6) which requires estimating the gender differentials in the scores across GAT, ACT and SGPA.

Research Question 1:

To answer research question one, a bivariate Pearson correlation coefficient was used to explore the relationship between the student accumulative assessment score at the end of high school education (SGPA) and the accumulative rate in the College of Education (UGPA) at the end of year 2014 at King Faisal University.

The results in Appendix F indicate a statistically significant relationship between the general High School GPA (SGPA) and the College of Education GPA (UGPA) at King Faisal University. The correlation coefficient is significantly different from zero (r=0.562, p<0.01). For a better interpretation of the correlation coefficient, Field (2013) advises that the coefficient should be squared to obtain a coefficient of determination which measures the percentage variance in one variable explained by the other. In this case, the coefficient of determination is obtained by squaring 0.562 and this equals 0.31. To this end, SGPA explains 31% of the UGPA, while the remaining 69% of variance in UGPA is explained by variables other than SGPA.

Research Question 2:

To answer question two, a Spearman's correlation analysis was undertaken to estimate the strength of relationship between the students' General Aptitude Test (GAT) scores at the end of High School education and the UGPA at the end of year 2014 at King Faisal University. This technique was chosen given that initial tests revealed the presence of a non-linear relationship between GAT and UGPA.

The results set out in Appendix G indicate that there is a statistically significant relationship between the General Aptitude Test (GAT) and UGPA *(r=0.32).* Though the size of the correlation coefficient signifies a medium effect relationship (see Field, 2013, pg. 267), it is

highly significant (p<0.001). The coefficient of determination (R square) is 0.10, implying that GAT explains 10% of the variation in UGPA, while the remainder, 90%, is explained by variables other than GAT.

Research Question 3:

To answer question 3, a Spearman's correlation analysis was undertaken to estimate the strength of the relationship between students' scores on the Achievement Test (ACT) at the end of High School education and the UGPA at King Faisal University. A decision to employ the Spearman's correlation technique was informed by the preliminary tests, which indicated a non-linear relationship between ACT and UGPA, and, hence, required a non-parametric technique.

From Appendix H, there can be seen a weak correlation between the Achievement Test score (ACT) and university GPA (r = 0.268). Though the relationship between the two was found to be weak, it is statistically highly significant (*p<0.001*). The coefficient of determination (R square) for this relationship is 0.072, implying that ACT explains about 7.2% of the variance in UGPA, while the remaining 92.8% is explained by variables other than ACT. Therefore, inasmuch as the ACT score is significantly associated with UGPA, this influence is minimal, as evidenced by a low coefficient of determination. The weak correlation found by the analysis is due to the non-linear relationship between the ACT score and UGPA.

Research Question 4:

Question 4 required establishing the joint effects of SGPA, GAT, and ACT on UGPA. A multiple linear ordinary least squares regression analysis is the most suitable analytical

technique for answering this question. <u>The general format of a multiple regression model is</u> <u>as follows:</u>

$$\underline{UGPA = b_0 + b_1 SGPA + b_2 GAT + b_3 ACT + \varepsilon}$$ ………………………………….Equation 1

<u>Where:</u>

<u>UGPA is the predicted variable, University GPA.</u>

<u>b_0 is the value of UGPA on the assumption that no other variable predicts it. In other words,</u> <u>this is the mean of the group when all the predictors are zero.</u>

<u>SGPA measures the cumulative score at a generally high level.</u>

<u>GAT is the student score from the General Aptitude Test.</u>

<u>ACT is the student score from the Achievement Test.</u>

<u>b_1-b_3 represent the relationships of the three independent variables, SGPA, GAT, and ACT,</u> <u>on UGPA.</u>

<u>ε is the error associated with the estimation of the model.</u>

In estimating the regression model, a stepwise method is used to enter the predictors. A stepwise method entails adding one predictor variable at a time and estimating the regression model. The choice to use this approach is based on the need to monitor changes such as those in the model fit that may occur on adding a predictor variable into the model.

From Appendix I, the first model which had SGPA as the only predictor accounts for 31.5% of the variation in UGPA. When GAT is added to model 1, the resulting model is model 2. There is a marginal increase in R square from 31.5% to 34.0%. This represents only a 2.5% increase. In model 3, ACT is added to the previous two predictors and there is no change in

R square, implying that ACT does not explain any variation in UGPA if estimated simultaneously with the other two variables.

From Appendix J, there is evidence to suggest that all three models significantly improve my ability to predict UGPA compared to a model without predictors. For instance, the first model is highly significant in predicting UGPA [$F_{(1,691)} = 318.3$, $p<.001$]. The second model is also significantly different from an empty model [$F_{(2,690)} = 177.744$, $p<.001$], and so is model three [$F_{(3,689)} = 118$, $p<.001$].

From Appendix K, the results for the final model (model 3) are interpreted. Overall, two out of the three predictor variables explain UGPA. These are SGPA and GAT. ACT does not significantly influence UGPA. Specifically, there is a statistically significant positive relationship between students' High School GPA scores and the university GPA score such that a one standard deviation increase in SGPA leads to a 0.512 standard deviation increase in UGPA ($\beta=0.512$, $p<.001$). This therefore means that students who perform well in their end of High School cumulative assessment (SGPA) are more likely to perform better in the university cumulative assessment (UGPA). This is expected given that students with higher prior learning attainment tend to achieve higher at subsequent academic levels.

The General Aptitude Test (GAT) score is statistically significant in predicting university performance (UGPA) such that a one standard deviation increase in GAT leads to 0.163 standard deviation increase in UGPA ($\beta = 0.163$, $p<.001$). This implies that students who perform better in the General Aptitude Test are more likely to perform better in the university cumulative assessment (UGPA). This conforms to earlier findings that learners with greater aptitudes are more likely to perform better in academics (see Thompson & Zamboanga, 2004).

Finally, though the ACT is positively associated with UGPA, this relationship is not statistically significant ($\beta = 0.006$, $p<.001$). This confirms the preliminary analyses involving

the correlation and assumption testing where it was found that ACT was weakly associated with UGPA. This result arises out of the realisation that ACT is not linearly associated with UGPA. Possibly, this could be a matter of significant non-alignment between the contents of the two assessments (ACT and UGPA) such that knowing of one does not support the passing of the other.

Research Question 5:

Research question 5 requires the identification of the most influential admission standard. The standardised coefficients (Beta) are used to compare the magnitude of influence of the independent variables on the dependent variable. Standardised coefficients are on the same scale and are, therefore, comparable. The bigger the size of the Beta coefficient, the greater the influence of or the ability of an independent variable to predict the dependent variable.

From Appendix K, it is evident that the SGPA is the most important factor in predicting the performance of students at the university (UGPA). Its Beta coefficient is more than three times that of GAT and about nine times the effect of ACT on UGPA. This implies that SGPA should be given a much greater weighting in the university admissions criteria, given such an influence. The second variable influential in predicting UGPA is GAT, while the variable without any significant influence on the university's admission criteria is ACT. This provides evidence for the need to reform the university admissions criteria by emphasising the variables that have a significant influence on UGPA and to do away with those that do not sufficiently influence UGPA.

Research question 6:

Research question 6 requires the estimation of the gender differences in GAT, ACT, and SGPA scores. Furthermore, the same research question requires estimation of the difference in GAT, ACT and SGPA scores according to whether or not one has changed major. To estimate the required gender differences, an independent samples t-test is undertaken. This technique is suitable for estimating differences between two groups with reference to scores in the GAT, ACT and SGPA. On the other hand, to estimate the differences in GAT, ACT and SGPA by change of major, an independent samples t-test is also undertaken. This is because the grouping variable manifests in two groups i.e. whether or not one has changed major.

It is important to note that, given that Achievement Tests (ACTs) were only recorded for female learners, it is not possible to conduct a t-test for them because there is only one group.

Appendix L indicates that there are differences in the SGPA and GAT on the basis of gender. For instance, female students performed better than male students in the SGPA (Male mean score = 309, Female mean score = 384). Similarly, the same trends in differences are manifest in the GAT scores. It is, however, not possible to conclude whether these differences are statistically significant by only basing them on descriptive statistics. Therefore, the next analysis is undertaken to determine statistically whether the gender differences highlighted by the descriptive statistics are statistically significant or not. This is done using an independent samples t-test.

The independent samples t-test relies mainly on a key assumption of equality of variances within each of the two groups. The Levene's Test is used to test for equality of variances.

The findings in Appendix N, indicate that the assumption of equality of variance (homogeneity of variance) is upheld in the GAT group scores but is violated by the SGPA group scores, given that the Levene's test for SGPA is statistically significant. The Welch's

robust test of equality of variance also confirmed the significant differences in the groups (see Appendix M). This implies that adjusted t-statistics and degrees of freedom are to be used.

From Appendix N, there are significant differences in the GAT and SGPA mean scores between male and female students. For instance, for SGPA, girls attained a mean score of 96.00%, while boys attained 85.15%, and the mean difference is statistically significant $[t(564.25) = -29.97, p<0.001)]$. Similarly, in the GAT test, girls performed better than boys (girls mean = 77.81%, boys mean = 72.33%) and the mean difference is significant $[t(691.00) = -10.00 \ p<0.001)]$. Therefore, the stated null hypothesis is rejected, and it can be concluded that there are statistically significant differences in the scores for GAT and SGPA by gender. Therefore, gender is a significant predictor of these two variables.

Appendix O shows that the two groups (No Change Major and Yes Change Major) show differences in mean scores across SGPA, GAT and ACT. The group means for the "No Change Major" are higher than their counterparts (Yes Change Major) across GAT and ACT. On the other hand, however, the "Yes Change Major" group scored higher means on SGPA than its counterpart did. However, it is not possible to tell whether the mean differences highlighted by the descriptive statistics are statistically significant. This requires a parametric analysis to obtain an answer. To this end, an independent samples t-test is performed. The key assumption of equality of variance is tested before embarking on the t-test.

The Levene's test for homogeneity of variance indicates that the assumption is upheld in the scores for GAT and ACT, while the same assumption is violated by SGPA. This implies that the estimated parameters for SGPA are adjusted for the violation.

From the results in Appendix Q, it is clear that the means of the two groups significantly differed across all admission criteria (SGPA, GAT, and ACT). Specifically, those that

changed their major had a higher SGPA mean compared to those that did not change major (change major mean = 94%, no change major mean = 90.9), and the mean difference is statistically significant [t(62.12) = -3.44, p<0.001]. On the other hand, for GAT, those that did not change major had higher mean scores compared to those that changed major (change major mean = 71.2%, no change major mean = 75.7%) and this mean difference is statistically significant [t(691) = 4.1, p<0.001]. Similarly, for the Achievement Test (ACT), those that did not change major had higher mean scores compared to those that changed major (change major mean = 69.9%, no change major mean = 81.4%) and this mean difference is statistically significant [t(382) = 10, p<0.001]. We can therefore conclude that there are statistically significant variations in the scores for the SGPA, GAT, and ACT by change of major. This implies that change of major predicts these three variables.

6.4.1   Summary of findings

From the above statistical analysis, it can be concluded that:

1. High School degree significantly predicts university GPA such that students who perform well at their High School are more likely to achieve a higher GPA in the Education department at King Faisal University.

2. The General Aptitude Test significantly predicts university GPA such that students with higher General Aptitude Test scores are more likely to score higher in the Education GPA.

3. There is a significant relationship between Achievement Test scores and university GPA such that students who join university after scoring highly in the end of High School Achievement Test are more likely to achieve higher GPA scores.

4. Jointly, only two (High School degree and General Aptitude Test) of the three independent variables predict university GPA while the Achievement Test score does not predict university GPA if all variables are concurrently estimated.

5. Using the standardised coefficients, it is clear that the High School degree score is the most influential factor in predicting university GPA while the Achievement Test score is the least influential in predicting university GPA.

6. There are significant gender differences in the High School degree scores and General Aptitude Test scores, with female students scoring higher than male students. On the one hand, students who had changed major had higher scores for the High School degree compared to those who did not change major. On the other hand, for the General Aptitude Test, those who did not change their major had higher mean scores on the General Aptitude Test.

6.5   Findings from interviews

Research questions seven and eight were answered using the semi-structured interview approach. As explained in the methodology chapter, the researcher used semi-structured interviews with open-ended-questions. The researcher interviewed eight academics who were then teaching a number of Education courses to students across a range of year groups. These modules covered the following areas within the degree: curricula and teaching methods, teaching strategies, educational psychology, principles of educational research, special education methods, health and fitness, sports science, and so forth. These participants included Lecturers, Professors, Assistant Professors and the Dean of the College of Education. Of the total number, two participants had studied in the USA, three had studied in the UK, and three had studied in SA. The interviewed participants had wide-ranging teaching careers with time in Higher Education ranging from eight to thirty years within the

Faculty of Education. Some lecturers had actually had direct personal experience of all the admission criteria currently used in the Faculty of Education as students themselves, while others finished their High School education before the establishment of the National Centre for Assessment (QIYAS). All respondents were men because of the difficulty of securing women's approval for interview given the social and gender constraints of Saudi Arabian culture.

Some of the interviewees in the study sample were members of the Centre for Assessment, either participating in the preparation of the QIYAS tests or being involved in the observation of students while they were taking the assessment tests. Therefore, these interviewees were considered as having considerable knowledge of the conditions of Saudi education in general and the methods of assessment used in the Faculty of Education and in High School in particular. All interviews took place during the academic year 2016-2017, and each interview lasted approximately 50 minutes.

Following the coding and inductive analysis procedures that were discussed in Chapter 4, seven main themes emerged from the semi-structured interviews: (1) Traditional methods continue, (2) Lack of variety of assessment, (3) The relationship between High School assessment and Education College assessment already exists, (4) Knowledge is priority, (5) Insufficient admission standards, (6) Changing subject, and, (7) Girls' performance. These themes are intended to respond to Research Questions 7 and 8 in this thesis. The themes are set out and explicated below:

6.5.1   Traditional Methods Continue!

Participants saw traditional assessment as referring to conventional methods of testing which typically lead to a written document, such as a quiz or an exam (Quansah, 2018).

Standardised tests, achievement tests, tests given to students by teachers or lecturers to measure how much the students have learned were also regarded as examples of traditional assessment (Ibid.). According to these participants, the most commonly used traditional assessment tools included multiple-choice, true/false, short answers, and writing essays, which focus on the lower levels of Bloom's Taxonomy (Dikli, 2003; Zimbicki, 2007) and involve relatively lower order cognitive skills.

This theme refers to the ways in which interviewers talked about the persisting place of traditional methods of assessment in their university. All participants, regardless of academic rank or amount of teaching experience in Higher Education, made reference to the fact that traditional methods of assessment continued to be the most common in the teaching of modules within the College of Education. Comments from both Participants1 and 3 who had studied in the UK and from one of the junior Lecturers were examples of this reference to traditional assessment:

> *In the faculty, we mostly use the traditional methods of assessment* (P 1).

> *In general, written tests were the most common method used in teaching at Education College (P 3).*

> *The types of tests that were prevalent in the testing and control rooms are traditional assessment either objective or essay-based* (P 5).

It was evident that these traditional methods had been in place for some time. However, some participants did mention that other forms of assessment were present but were infrequent. While including traditional methods in their modules, there was some preference for using other methods including presentations, portfolio, peer assessment and projects. Comments came principally from the more junior lecturers':

Of course, I use the traditional tests and the portfolio, which is the performance file, especially in the practical education module (P 2).

For example, the student portfolio, student presentations, and projects that I ask my students to prepare (P 6).

All interviewees gave their reasons for continuing to use more traditional methods of assessment in their modules. Rather than their educational benefits, the most common reason was the presence of very large numbers of students in the classroom that, in their view, made the choice to use assignments or tests easier to manage and ultimately to mark. This focus on practicality was certainly evident in the comment from an Assistant Professor and from one of the new Lecturers in the Education College.

Doctors and Lecturers usually opt for the testing system that would make the job of correction easy for them (P 1).

I still think that the traditional methods of assessment are the most convenient because they are in keeping with the aspirations and abilities of students in Saudi Arabia (P 8).

A further reason for continuing to use more traditional methods of assessment in modules, was a concern that undergraduate students might not have the appropriate abilities or 'skills' to demonstrate when being assessed by other methods. Comments from participants 6 and 7, who were the youngest members, of staff were examples of this.

Therefore, when students are asked for new methods that require personal effort, they may be surprised or may not have sufficient skills to prepare such activity or assessment (P 6).

The students in the High School do not have to give a presentation, submit a paper, or even engage in a group discussion … I think if they do this, it would be a massive

boost to their chances to gain a university place … In other words, they will be admitted into university having had the opportunity to experiment with these methods as a means of assessment (P 7).

One participant, who had studied in Britain, ,went as far to claim that the main reason for using traditional methods of evaluation was the poor capabilities of the teaching staff in the College of Education!

They stated; "*The College of Education, which is supposed to provide specialised instruction, and whose members are meant to have knowledge and adopt diverse approaches towards modern assessment methods, is in fact one of the weakest and less advantageous colleges in the use of assessment methods, which is also the case for the teaching process at KFU*" *(P 1).*

The same participant also directed some stern criticism towards the Faculty at the College:

"*To be honest with you, the traditional assessment process we have in place is a routine procedure that is only focused on finishing the course. There is not only a serious disinterest in assessment for its own sake, but also many of us lack the real skillset needed for carrying out the right kind of assessment!*"

This reference to a lack of assessment knowledge is consistent with several findings (e.g. Al-Amri, 2014; Al-Kareani, 2010; and Al-Otaibi, 2006), in which researchers highlighted the mediocre assessment practice among Faculty members in the Colleges of Education in Saudi Arabia in terms of developing their skills in the field of learning and assessing learners, and their application of educational practices in the delivery of modern materials. Such a deficit was seen as a contributing factor to weak outcomes within these Faculties. Traditional assessment strategies are no longer able to keep abreast with the knowledge explosion and the ICT revolution (Al-Maghdawi, 2018). From this perspective, it is necessary for the

Colleges of Education in Saudi universities in general and the Faculty of Education at KFU in particular to undertake a comprehensive review of their educational systems in general and the assessment system at the modular level. The interview data pointed to some criticism of existing assessment practices and that, in the main, junior lecturers and those with some higher education experience beyond Saudi Arabia were showing some enthusiasm to undertake reform. To that end, the College of Education should undertake a bold development process in order to employ modern evaluation strategies that have proven important in developing the capabilities and skills of learners, as explained in *Chapter Three* of this thesis. This work would also lead to revisions of regulations of the College of Education so that they comply with the requirements of the labour market and higher education policy in Saudi Arabia so that *Vision 2030* is explicitly accomplished. Such *Vision* seeks to develop curricula, teaching methods and assessment in a way that positively affects undergraduate student outputs and the labour market thereafter.

6.5.2   Lack of variety of assessment

As briefly referred to in the previous theme, this limited variety of assessment used by participants to assess student performance emerged from the interview data. Although participants referred to the place and use of more traditional modes of assessment within the College, there were some genuine but infrequent attempts to use different or alternative forms of assessment alongside what might be termed traditional forms of assessment. The choice to use alternative forms of assessment was more common among Lecturers who had studied overseas. To illustrate, one of the participants claimed to have started to use social media as a tool to assess their students.

I have recently started to use WhatsApp as a means of communication and evaluation (P 2).

There was occasional mention by Lecturers of seeking to place assessment in the hands of their students - the Dean referred to their use of one type of alternative assessment characteristic of Assessment for Learning.

I also asked all the students to undertake peer assessment by commenting on the performance of the student presenting in front of them (P 4).

However, it was evident in interviews that these trials could be seen as personal efforts at experimenting with alternative forms of assessment and therefore appeared not to be 'official' attempts undertaken with the support of the Faculty. This small trend within the College of Education thus appeared to go against the requirement to use a mixed package of traditional modes of assessment, which, from the data, is understood as essays, true/false and multiple-choice questions. In fact, there was occasional reference to the restrictive 'official' Education College directive about the limits of assessment use which applied to all academics within the Education College, as the Dean indicated:

Therefore, a decision has been made in the Faculty Board in all departments that a member of the teaching staff at the Faculty of Education must plan at least 50% of the test in essay format (P 4)

A few of the academics who had studied overseas were therefore quite anxious about using such alternative assessments, for fear of 'being seen as thinking outside the box'. By this, such academics saw themselves as the only colleague within the Education College who used these methods of assessment. One Associate Professor who had studied in the US shared their view:

Because if I rely entirely on modern assessment methods that recent theories seek to publicise or adopt, perhaps I would look like the odd one out among my colleagues (P 6).

Furthermore, these lecturers faced further challenges that prevented them from continuing to apply these types of alternative assessments. Examples of these referred to in interviews included: undergraduate students choosing to withdraw from courses where alternative assessments were being used, a lack of prior experience among learners with these forms of assessment, and an apparent absence of a system within the faculty that supports innovation among lecturers in their application of these types of assessment, even in the form of a pilot attempt.

Participants referred to the challenge of changing the attitudes of learners who seemed only to be willing to accept or adapt to a particular type of assessment (traditional assessment) in complete rejection of other types. There was a shared view that alternative assessment should start from the early stages of education which emerged. Comments from participant 6 who had studied abroad and from lecturer 7 who had studied in Saudi Arabia are examples of this:

> *If we want to adopt some alternative assessments methods of education, we must start thinking about implementation from the primary school (P 6).*

> *I mean we might start by introducing the various methods of assessment into the High School so that the students are well-prepared to face any problems and are confident enough to take on their studies at the university (P 7).*

The Lecturers who tended to support alternative assessment saw it as appropriate in addressing the improvement of learning outcomes which could address the needs of the labour market in Saudi Arabia.

> *You may know that there is a gap between the outputs of higher education and the needs of the Saudi labour market, which is clear and has been frequently echoed. Teaching methods and the rigid and old-fashioned evaluation methods, whether in*

174

general education schools or at universities, may pose as an underlying reason for this gap. Therefore, we must keep pace with the global developments in higher education as there are great efforts exerted in such education means and methods for the advancement and better preparation of graduates (P 4)

An interesting pattern emerged again from the responses in that it seemed that the majority of those who did complete their postgraduate/doctoral studies outside of Saudi Arabia made attempts to introduce alternative assessment approaches, such as WhatsApp, oral presentations, and peer assessment in their assessment of learners' achievement. While this is largely speculation, a reason could be that they or their children have been exposed to those types of assessment while abroad and that they personally found a range of benefits to be associated with these types of alternative assessment and, thus, tried to replicate what they believed to be effective with learners at the Faculty of Education at KFU.

However, the other participants were opposed to this and these were from more senior colleagues who may be wishing to retain the traditional approaches and are perhaps reluctant to change. Comment from participant (8) was examples of this view

The traditional methods of assessment are the most convenient because they are in keeping with the aspirations and abilities of students in Saudi Arabia compared with alternative assessment (P 8).

6.5.3 The relationship between High School assessment and Education College assessment

This theme provides an account of the claimed views among interviewees on the relationship between the methods of assessment used in High School and those which are employed in the College of Education. Thus, this theme refers to the assessment methods that are used in

High School and which are being used in the Education College at KFU, which are based on writing exams and essays.

Most of the interviewees were of the opinion that the assessment methods currently being used were basically aligned in 'assessment form'. By that it can be understood that there was no clear differences between the assessment methods found in the respective educational stages. This position is not entirely positive, but it may violate the Ministry of Higher Education's policy of developing learners who are proficient in necessary labour market skills. Perhaps the most important issue referred to by the Saudi Ministry of Higher Education is the promotion of teaching methods that places the learner at the centre of the learning process and not the teacher, in addition to focusing on building skills, refining personality, cultivating confidence and honing the spirit of creativity. This aspiration requires a comprehensive review of the Saudi universities' systems and goals, as well as looking at the evaluation methods used, and ways to communicate information to learners so that they are part of acquiring knowledge and skills. The data in this thesis suggest some potential for change given the views of some lecturers concerning alternative assessment. However, some of the most junior lecturers with the most recent experience of being assessed at the end of high school experience confirm the position to be reformed. For example, participant 2 who had finished his High School before QIYAS was established and Participant 8 who had undergone the QIYAS tests confirmed that: *'going back to your question, we do not differ much from the High School in the methods of assessment' (P 2).* While another participant added;

There are existing links between student assessment methods such as traditional tests (P 8).

Participant 5, who had studied in Saudi Arabia, also confirmed this view on the alignment between the methods of assessment used in High School and in the Education College.

However, he went on to be somewhat critical of these assessment methods on the grounds that they did not measure higher levels of student knowledge, skills and understanding: *The assessment methods used in the High School are the same in the Faculty of Education. They are still based on the measurement of the minimum skills. This is because the minimum skills are easy to measure (P 5).*

The same participant also added that:

> *As the researcher in this study, perhaps what you mentioned in that the high school diploma is the most closely related to the success rate in the college is an indication of the fact that the assessment methods in the high school are similar to the methods used in the schools of the college of education.*

In order to develop a greater awareness of other assessment methods, both in High School and in the Faculty of Education, an Assistant Professor saw the need to pay closer attention to the content of assessment that was currently part of pre-service teacher training in Saudi Arabia:

However, the key problem is in the area of preparing and training teachers on the latest approaches [to assessment] and teaching styles in education…..In reality, there is no link in my mind right now except to look after the students of the faculties of education. This can be done by training them in modern methods and tools in the field of evaluation (P 3).

Some revision to this aspect of the teacher training curriculum was seen as crucial, given that most Education students will ultimately become teachers in Saudi's public and private elementary and secondary schools.

It should be pointed out here that, in the minds of participants, the responsibility lies at the feet of the College of Education in terms of addressing the weaknesses of training and

qualifications related to alternative assessment methods, especially for students who are destined for a teaching career. This is in addition to monitoring the students' applications during field application in the last term of the bachelor's degree. Similarly, the College of Education should conduct training courses on alternative assessment methods for its lecturers in order to potentially facilitate undergraduate students' acceptance of the various alternative assessment methods and to ensure fairness in the overall evaluation of students, as not all students prefer the traditional assessment approach.

6.5.4   Assessment of knowledge as priority

This theme details the limited attention paid to 'knowledge' as measured by the current assessment methods in High School as well as those typically used by the College of Education at KFU. The interviews tended to foreground the first three level of Bloom's Taxonomy as a priority when measuring a student's performance in the Education College at KFU. The first three levels of Bloom's Taxonomy link with skills (remembering, understanding and applying) that tend to deal with lower-order thinking and that are seen as important in laying the foundation for understanding, as opposed to the last three which employ higher-order thinking skills (analysing, evaluation and creating) (see Hopper, 2012).

All participants noted that only minimum levels of knowledge were currently being measured by assessment methods used both at the High School level and in the Faculty of Education and that this did not always align with the assessment expectations at the university, official or otherwise. This is certainly not in keeping with the global direction of higher education, which is based in most developed countries on the marketing policy, which tries to connect universities to the job market (Dill, 2003; Molesworth et al., 2011; Tomlinson, 2018 and Alshamlan & Al-Fawzan, 2017). As such, the job market needs graduates who do not only have knowledge, but also skills and mental flexibility. Participants (3) and (8) shared their perspectives on this narrow attention to 'skills':

We have qualitative questions that evaluate the higher aspects of the student's mental processes, such as critical thinking and the ability to be creative. Unfortunately, these abilities are not measured in both High School and also in our College (P 3).

The assessment methods used in the High School are the same in the Faculty of Education. They are still based on the measurement of the minimum skills (P 8).

One of the Assistant Professors, who cultivated an interest in Education assessment, indicated that the current assessment methods tend to focus only on memorisation and understanding, referring to a study conducted by High School students. The aim of the study was to analyse the tests offered to students in High School in the city of Alahsa. He concluded that:

I did carry out a study about this topic in the past. The following results came out with 80% remembering and understanding, while application stood at 20% (P 1).

Assistant Professors who studied outside Saudi Arabia offered a reason for the apparent lack of attention to measuring higher levels of knowledge:

The reason could be that educational curricula do not support the development of these capabilities (P 3).

While one of the lecturers mentioned that:

The reason that some colleagues focus on knowledge may be due to their expectation that what suits their past time in higher education, which is knowledge, may fit this time with its development, and this is undoubtedly not entirely correct (P 7).

In addition to an evident narrowness of curriculum, Participant (6) suggested that diversification in assessment methods was the key to achieving a much broader measurement

and display of different levels of knowledge and skills, which current modes of assessment tended to ignore:

The measurement and assessment methods must vary in order to better serve this profession in the future. Also, the variety serves the graduates in terms of their ability to face the audience and to speak, as well as their ability to recap and prepare lessons in a professional manner (P 6).

Given its major role for those wanting to work in public or Higher Education, perhaps one of the most important challenges facing the Faculty of Education at KFU is to broaden its learning outcomes and expose undergraduates to different methods of assessment that can evidence gains in knowledge and understanding. Reinforcing this viewpoint is the fact that 35% of Saudi citizens are, in fact, students in general and Higher Education. The *Vision* of the College of Education would be best constructed to explicitly align with *Vision 2030*. Accordingly, other than seeking to build 'softer' skills and creativity, one of the most important objectives of education evident within *Vision 2030* is to promote teaching methods and pedagogies that place the student at the centre of learning rather than this being traditionally the locus of the teacher.

6.5.5   Insufficient admission standards

A common view held by participants was that the three admission standards currently used in the Faculty of Education (High School Test, Abilities Test and Achievement Test) were not sufficient as criteria for judging student acceptance into or rejection from the Faculty of Education at KFU. However, the majority of participants agreed that academic performance at High School still maintained its rightful position as an acceptable criterion for admission to and potential measure of future performance within the Faculty of Education and in

university more generally. Comments from participants (5) and (6) were indicative of this view:

> *In my personal opinion, High School has the highest predictor of success of students at University (P 5).*

> *I attribute this to the fact that High School is still a strong factor in determining High School students' performance and in accepting them in the Faculty of Education (P 6).*

While such support for the above criteria was evident, it was apparent that some participants wished to see more robust confirmation of their suitability for use in the current Saudi system. Their comments referred specifically to the QIYAS Centre tests and the need for the Centre to report more evidence to support appropriate use of this testing at High School. Participants 3, 5, and 6 in particular were quite sceptical about this lack of evidence, and felt that further research was required:

> *There has to be a study to support the answer for this. I haven't come across any studies that have tackled this aspect about QIYAS tests (P 3).*

> *The reason is simple in that there is no published studies yet to prove the success of these tests whether ability test or Achievement Test (P 5).*

> *The argument for or against the Achievement Test, ability test and the degree of credibility or suitability of students requires many researches so that one can ascertain their worth (P 6).*

A couple of other respondents, the Dean of the Education College and a member of a QIYAS centre, went as far as to completely question the tests of the QIYAS Centre and instead

supported some reconsideration of the whole assessment process due to issues regarding its current standing and imprecision.

As for the tests of a QIYAS centre, there are actually questions of achievement and abilities, but they need to be questioned in terms of their reliability ...To be honest, I was personally involved in QIYAS and found some issues of inaccuracy in the criteria put forward in some of the tests, but I cannot tell you which tests these were. This means that there is a great difference in how you and I interpret these criteria (P 2).

So, if I can explain it better, the issue of the lack of credibility of the tests, which is achievement and abilities tests, in my opinion, has made the High School more pressing (P 4).

One of the supporters of the High School as the criterion of admission added a further shortcoming of the criteria of the QIYAS Centre tests, which they regarded as more of a test of 'guessing' as opposed to a rigorous test of knowledge and application.

The tests in the QIYAS Centre are basically multiple-choice tests, so they also have a 25% guess rate (P 5).

Despite all these criticisms, there was, however, an accepted recognition and admission that these tests were required by Saudi law as the necessary standards for entry, as mentioned by the Dean of the Education College.

But the only codified test in the country is the QIYAS tests, which cannot be dispensed with, in fact (P 4).

Although somewhat resigned to the fact that QIYAS tests were still mandatory, most participants tended to support the addition of further criteria given that the current criteria in their view were not seen as sufficient to appropriately judge both the admission of students

into the Faculty of Education and offer a prediction of their future performance. Participants (4) and (6) supported this point.

These tests still need other indicators for our judgment to be fair (P 4).

From my point of view, I don't think those standards are enough to give indicators about predicting performance or judge on the students to accept them at the university (P 6).

Similarly, a few combinations of Lecturers suggested adding a new admission criterion:

I think that ... if you notice the standards that are currently used in the college, they all depend on the standard tests, whether in high school or the tests of the Qiyas National Centre for Assessment. As such, we cannot identify the skill aspect, nor can we determine the extent of the applicant's ability to rise up to challenging situations or speak in front of an audience, as in the interview, is not possible according to the current standards because it is briefly a paper and pen examination system (P 2).

There were other advocates of moving to an interview as one entry criterion; *"I believe that the interview is a strong criterion when assessing a person" (P 6).*

In addition, some participants suggested designing a 'test' specifically for the Faculty of Education and that this test would serve as an additional admission standard. This test includes the assessment of some of the required skills for the students admitted into the Faculty of Education, such as presentation skills, knowledge related to educational theories, especially in relation to education, and the extent to which students are familiar with the various assessment methods and their acceptance of such methods. The Dean said that:

I mean that after the High School, students should be given an initial test to identify if they are suitable for the course and to check their abilities (P 4).

Finally, one of the participants argued that the current standards are sufficient; however, he suggested that the percentage of each admission criterion should be amended.

There is a prevailing view that the relative weight allocated to the test of abilities and achievement is very high. In my meeting with many parents and interested people, the secondary test assumes the largest relative weight in admission to the Faculty of Education (P 3).

6.5.6   Changing their subject/major

Changing academic subject or major tends to occur in the first year of university study. In this study, it refers to changing the subject by joining the Faculty of Education at King Faisal University from another College inside KFU during the first academic year.

It was mentioned that one of the reasons for changing subject may be because of initial parents' opinions and their influence on the subject choice of their son. A comment from an Associate Professor is an example of this.

Sometimes the student choose for a subject that he or she doesn't want because of parental pressure or the need for it in the labour market (P 2).

Another reason could be the lack of interest of the learner in the previous subject. A comment from the Dean exemplifies this.

They may feel psychologically at ease and perform much better because they do not like the subject they're enrolled into (P 4).

Half of the respondents reported that changing subject could contribute to raising the Grade Point Average in the Faculty of Education. What the Assistant Professor and Associate Professor said were examples of this:

I have come across many cases when the student changes the subject matter and does much better in the new subject... Change is often positive I would say (P 2).

We expect that when the student selects the specialisation he would like to pursue, this contributes to his achievement of a high mark in the faculty (P 3).

Some interviewees gave an example from their personal experiences to illustrate their points. Comments from the Dean and from an Assistant Professor were good examples of this:

It happened to me personally. I was a Maths undergraduate, I changed to the giftedness specialisation. I found myself in the field of caring for the gifted (P 1).

When some students change subject from one subject to another, they may feel psychologically at ease and perform much better because they do not like the subject they're enrolled into. This happened to me personally (P 4).

The other half of participants, however, mentioned that the impact of subject change depends on the personality of the learner. It depends on the students and what career path they want to follow.

Therefore, it is possible that the effect will come from the student's own circumstances and his or her performance (P 5).

The impact of changing subject may be a catalyst for success, as mentioned by participant 6 who was previously a president of the curriculum and teaching methods.

If the student is strongly motivated in his or her new specialisation, I don't think this will negatively affect them. But if there is a great desire and the grade is pretty good, then maybe the student will maintain his level or even raise it (P 6).

6.5.7   Females do better

An original element of this thesis involves paying attention to the performance of female learners in Saudi Higher Education and reporting this comparatively to that of male undergraduates. Female students' achievement refers to the cumulative average at the end of the final year of study in the Faculty of Education, which is sometimes called the Grade Point Average of the bachelor's degree. Some of the outcomes following the statistical analysis developed through the SPSS programme which accounted for the achievement of male and female students were shared with the participants during the interviews. The purpose of this was to gather the perspective of the male academics concerning the performance of female students in their academic achievement compared to that of male students. Most of the respondents provided their reasons for the high achievement level of female students. It was evident that the reasons put forward tended to align with stereotypical taken-for-granted assumptions concerning the place and role of women in Saudi Arabia. Participants 1, 2 and 6, who had studied abroad in the UK and US exemplified this:

> *This could be because of the status of women in the Saudi society. The fact is that girls spend most of their time at home, which helps them to have a high achievement (P 1).*

> *As well as the characteristic of Saudi Arabian girls who are usually provided with the right conditions to make them succeed and excel (P 2).*

> *The reasons may be because the girls in this country have a special place and most of their time is spent at home as they only go out except for necessity (P 6).*

Occasional mention was made of the historical lack of work opportunities for girls.

There is also the cultural aspect of the Saudi society, in which jobs are available to boys more than girls … In other words, girls' employment opportunities are limited (P1).

However, the reason for this increase in the achievement of female students compared to students may be the inferior view of women in some regions of Saudi Arabia, as inferred by the subject who studied abroad, in particular highlighting the motivation among some women to challenge taken for granted views concerning women in the workplace. According to this interviewee:

To be honest with you, I studied in America about six years. The one thing you notice here is that we may have a different culture in some of our regions … For example, some regions do not allow women to engage in any field of work even if this is readily available to them … For this reason, the student seeks to prove that she is qualified for that type of work by striving for gain high grades for her own self-esteem, as well as to prove wrong those with negative attitude or inferior view of her (P 6).

A few respondents, however, attributed the reason for this high achievement to the fact that female teachers in High Schools were possibly more effective in the classroom than male teachers in High Schools for boys. A comment from Participant (2), who had studied in the UK, was indicative of this:

That's why the High School test results confirm the high level of girls' performance because the female teachers are better than the male teachers in terms of finding the motivation and commitment to the policies (P 2).

Others pointed out that, in their opinion, girls were often looking to outperform others and be the best because of the intense competition to secure employment that this can sometimes

lead to competing with their peers. Comments from participants (3), (5) and (6), two of whom had studied overseas during their PhD Journey, were examples of this view:

It could be that the level of positive anxiety among girls is higher than that of boys (P 3).

Jealousy could be another factor and is relatively high among girls (P 5).

Girls can also be sensitive to criticism from others, and hence they are usually trying to do especially well in their study (P 6).

Thus, these comments from the lecturers on female educational achievement seem to be aligned with *Vision 2030* in terms of the changing cultural and social lives of girls and women in KSA.

This section was intended to answer research questions seven and eight using the semi-structured interview approach. The researcher interviewed eight academics who, at the time, were then teaching in the Education College at KFU. These findings are now discussed in relation to the literature in the next chapter.

Chapter 7   : Discussion

7.1   Introduction

The aim of this chapter is to discuss the findings presented in Chapter Six in light of the literature reviewed in earlier chapters. This study investigated the predictive validity of the current admission standards which are applied at the College of Education at KFU in Saudi Arabia and explores which score among the current criteria offers the strongest potential contribution to students' future academic success. The current study will firstly be discussed in relation to the findings from the quantitative aspect of the thesis, which included the following research questions:

1.   Is there a statistically significant relationship between the student accumulative assessment score at the end of High School and the accumulative rate in the College of Education at the end of year 2014 at King Faisal University in Ahsaa, Saudi Arabia?

2.   Is there a statistically significant relationship between the student mark in the General Aptitude Test at the end of High School and the accumulative rate at the College of Education at the end of year 2014 at King Faisal University?

3.   Is there a statistically significant relationship between the student mark in the Achievement Test at the end of High School and the accumulative rate at the College of Education at King Faisal University?

4.   What is the degree of the following independent variables' ability of interpreting the accumulative rate of departments within the College of Education at King Faisal University in Ahsaa?

-    The accumulative rate in High School.

-    The marks from the General Aptitude Tests.

-       The marks from the Achievement Test.

5.   Which admission standards have the greatest ability to predict success at the College of Education at King Faisal University?

6.   Is there a statistically significant relationship between the score of the General Aptitude Test, Achievement Test, and High School percentage with respect to the variables of gender and change of major?

The discussion below is organised around the individual research questions.

7.2   Research question one

In this section, the chapter discusses findings from the first question in this study which is: 'Is there a statistically significant relationship between the student accumulative assessment score at the end of High School and the accumulative rate in the College of Education at the end of 2014 at King Faisal University in Ahsaa, Saudi Arabia?'

The results indicate a statistically significant relationship between students' accumulative rate in High School (SGPA) and the accumulative rate in the College of Education (UGPA) at King Faisal University (r=0.562, p<0.01). High School achievement explained approximately 31% of the total variance of the UGPA. This positive relationship suggests that students who performed well at their High School were more likely to achieve a higher GPA in the Education department at King Faisal University. This finding is similar to those of many studies that have indicated that High School percentages and Education College GPA scores are generally significant predictors of students' academic performance during their undergraduate studies (Alnahdi 2015; Alshehri, 2011; Alshumrani, 2007; Geiser & Studley, 2010; Kobrin et al., 2008; Murshid, 2013; Tesfa, 2013; Vulperhorst et al., 2018). In contrast, some researchers have found that the High School exit score was not a sufficient

predictor for success at university or that it had a weak relationship with the exiting GPA at university (Alkhattab 1989; Rahal 1987). Specifically, both Alkhattab's and Rahal's studies argue that the High School certificate alone is not a sufficient criterion for university admission due to its weak correlation with academic success and, therefore, these authors were also of the view that the certificate was an insufficient criterion for predicting success in university. As observed in the findings in this thesis, these two studies similarly recommended that there is a need to add further criteria for admission to university in addition to the High School percentage criterion as a prerequisite.

The findings from this thesis add to the existing evidence base which states that achievement outcomes from High School remain one of the best predictors for future academic success in the College of Education at KFU. Along those lines, Mathsiasen (1984) reviewed more than 60 studies and concluded that High School GPA was one of the best predictors of university performance. In the context of this thesis, Mathsiasen's findings were supported by Alshumrani (2007), who reported that High School percentage was seen as the best predictor of college success in Saudi Arabia. Closer examination of Alshumrani's research reveals this work was not undertaken in the context of students who were solely studying for an Education degree but included a number of Colleges at the University of Umm Al-Qura.

It is argued that the findings from the first question are consistent with the following logic. The High School GPA in Saudi Arabia is a cumulative measurement of the entire High School career of a student. It is calculated based on student scores gained in both the eleventh and twelfth grades. It stands to reason than this accumulation of achievement evidence for an individual pupil thus gives the High School exit standard strength as a predictor of academic success and likely supports the argument that the current assessment mechanisms are suitable and, therefore, this finding is informative for those in universities, in this instance in Education departments.

7.3   Research question two:

In this section, the research discusses the findings of the second question in this study which is 'Is there a statistically significant relationship between the student score in the General Aptitude Test at the end of High School and the accumulative rate at the College of Education at the end of year 2014 at King Faisal University?'

The findings indicate that there is a statistically significant relationship between the General Aptitude Test and the exit GPA of students within the Education College (r=0.32, p<0.01). The GAT explained 10% of the variation in UGPA. This positive relationship implies that students with higher General Aptitude Test scores were more likely to score higher in the Education GPA. This result concurs with the findings of other studies that also indicated the General Aptitude Test and the university final GPA scores were generally significant predictors of students' academic performance during their undergraduate studies (Al-Dossary, 2001; Alqatei & Alharbi, 2012; Alshehri, 2011; Alshumrani, 2007; Camara & Echternacht, 2000; Geiser & Studley, 2001;).

In contrast to this thesis, none of the previous studies appeared to have covered the Faculty of Education as a separate academic discipline. Although Alshumrani (2007) included the Faculty of Education in his research sample, the study, however, in contrast to this thesis, included only male students across all faculties of the Umm al-Qura University. Alshumrani's study was also conducted before girls were required to take the exit High School tests currently administered through the QIYAS Centre.

However, there are some studies where the findings revealed that the General Aptitude Test was not a sufficient predictor for future success at university (Geiser & Studley 2010; Murshid, 2013). In the current study, the relationship between the General Aptitude Test and success at the Education College was positive. There is indeed a statistically significant relationship, but the correlation coefficient is somewhat weak (r=0.32). This result may

provide some evidence that forms the basis for criticism regarding the validity of standardised tests in general and of the General Aptitude Test as a criterion for admission into Faculties of Education, in particular, in Saudi Arabia. According to Heilig and Darling-Hammond (2008) and Edmunds (2010), a number of critics disputing the rationality of standardised testing refer to the educational constraint of such testing and its failure to gauge the spirit of learning. Similarly, several viewpoints are in consensus regarding the significance of the class-time spent on the tests, which amounts to some kind of transaction cost approach to the argument and the likelihood that certain instructors/teachers 'teach to the test', as it is commonly known, thus likely skewing the outcomes (Shannon, 2008). Concerns over teaching to the test was evidenced in the responses of some lecturers interviewed for this thesis.

7.4   Research question three:

In this section the research discusses the findings of the third question in this study, which is: 'Is there a statistically significant relationship between the student mark in the Achievement Test at the end of High School and the accumulative rate at the College of Education at King Faisal University?'

The findings indicate that a statistically significant relationship exists between the Achievement Test at the end of High School and the exiting Education College GPA. The results indicate that students who join university after scoring highly in their end of High School Achievement Test are more likely to achieve better in their studies in Education ($r=0.27$, $p<0.01$). The ACT explained about 7.2% of the variance in UGPA. However, the relationship is still weak when this is compared to the other two standards referred to in research questions one and two. These findings correspond to much of the previous international research which also demonstrated that the Achievement Test and the Education

College GPA scores show significant relationships in terms of students' academic performance during their undergraduate studies (Alqatei & Alharbi 2012; Alshehri, 2011; Carty & Armstrong 2008; Murshid, 2013; Ramist, Lewis & McCamley-Jenkins 2001).

However, a review of the literature revealed one study which indicated that the Achievement Test and the Education College GPA scores were not significant predictors of students' academic performance during their undergraduate careers (Alsaif, 2005). Alsaif's (2005) study concluded that the cumulative rate for the foundation year was the most appropriate measure for predicting students' undergraduate achievement at the end of the first year. The cumulative rate for the foundation year has been interpreted as 35% of the total variation in achievement in the first year, followed by the High School rate, which improves the contrast percentage to 37%. As for admission tests (capacity tests), they only added 0.5%, which is a simple addition that does not affect the difference. Similarly, achievement has not affected such a discrepancy. Accordingly, Alsaif (2005) indicated that this result was indicative of the low predictive value of the admission tests used by the university in question. Alsaif (2005) also recommended that the admission criteria used at King Fahd University of Petroleum and Minerals should continue. It should be pointed out that the admission tests are designed by King Fahd University of Petroleum and Minerals and not by the QIYAS Centre because the sample chosen in the study had been determined prior to the establishment of the QIYAS Centre in Saudi Arabia.

In summary, these findings could lead to the conclusion that High School scores have a greater relationship with Education College GPA than the General Aptitude Test and the Achievement Test, considering separate criteria. The author of this thesis is of the opinion that there is a need to review the content of admission tests given their ineffectiveness in predicting academic achievements. This point will be addressed in the conclusion chapter.

7.5   Research questions four and five:

In this section, the findings from both the fourth and fifth questions in this study are discussed. These questions are 'What is the degree of the following independent variables' ability in interpreting the accumulative rate of department within the College of Education at King Faisal University?'

- The accumulative rate in High School.

- The marks of the General Aptitude Tests.

- The marks of the Achievement Test.

and 'Which admission standards have the greatest ability to predict success at the College of Education at King Faisal University?'

The findings in this study indicated that, when GAT is added to SGPA, there was a small increase in the R-square from 31.5% to 34.0%. This constitutes only a 2.5% increase. However, when ACT is added to the SGPA and GAT, there is no change in the R-square. This means that the ACT does not explain any variation in UGPA if estimated at the same time as the other two variables.

There is a statistically significant positive relationship between students' High School GPA score, the GAT, and the university GPA score such that a one standard deviation increase in SGPA leads to a 0.512 standard deviation increase in UGPA (β=0.512, p<.001), such that a one standard deviation increase in GAT leads to a 0.163 standard deviation increase in UGPA (β=0.163, p<.001). However, while the ACT is positively associated with UGPA, this relationship is not statistically significant (β=0.06, p<.001). This is in line with the earlier findings of the correlation and assumption testing where it was found that ACT was weakly associated with UGPA. To summarise, the accumulative rate in High School and the General

Aptitude Test are the most influential admission standards in predicting university GPA within the Education department investigated in this thesis.

The results for question five showed that the High School accumulative rate is the most important factor in predicting the performance of students in the Education College, followed by General Aptitude and then the Achievement Test. However, the Beta coefficient for the High School result is more than three times that of the GAT and about nine times the effect of the ACT on UGPA (0.512, 0.163 and 0.006 respectively). These findings indicate that using both the High School GPA and the General Aptitude Test could potentially help admissions employees within universities to make more accurate predictions and more appropriate admission decisions than using High School scores alone. These findings are similar to much of the previous research which found that the High School GPA explains more of the total variance in college GPA than standardised tests (Albanai, Balhadi & Alkhawli, 2004; Almiqdadi, 2010; Almokhlafi, 2001; Alshehri, 2011; Alshumrani, 2007; Darwaza, 1987; Murshid, 2013; Makransky et al., 2016).

In this thesis, the High School rate was the strongest indicator of success when the standard is the cumulative average in the College of Education. This finding is quite likely explained by the fact that the teaching methods used in the assessment of students in High Schools and Colleges of Education are very much the same. In other words, teaching and assessment in the Faculty of Education is an extension of teaching and assessment in High Schools. This calls into question whether there is any difference in the skills or knowledge that are measured by the methods of assessment or teaching methods used in the College of Education for the secondary stage (Alnahdi, 2015; Alqatei & Alharbi, 2012). Since the similarities between High School and university education outweigh the differences, it is only reasonable to assume that the cumulative rate of High School is the most accurate indicator of university success. As such, the methods of learning and assessment used in

universities are assumed to be different, deeper and more relevant to the labour market than those used in High Schools.

Many studies have also discussed that standardised tests more accurately predict academic success in colleges than the High School percentage does (Al-Hattami, 2012; Alqatei & Alharbi, 2012; Camara & Echternacht, 2000; Kobrin et al., 2008; Morgan, 1990; Shaw et al., 2012;). Other studies have indicated that standardised test scores add little information to prediction equations (Alshehri, 2011; Geiser & Studley, 2010; Murshid, 2013). These results therefore require researchers to examine other types of validity evidence (i.e., content and internal consistency) for standardised tests (Al-Hattami, 2012).

In the current study, there was a weak correlation between the cumulative student rate in the Faculty of Education and the Achievement Test, which may be ascribed to the fact that the Achievement Test measures different abilities and skills that may be overlooked in High School and university tests. This weakness in the predictive value of the Test of Achievement from the QIYAS Centre may require further development and necessary adjustments to be made to their tests to increase their predictive power. This is considered in the conclusion of this thesis.

Given the many criticisms of standardised tests mentioned in Chapter Three of this thesis (Ricketts, 2010; Smith & Fey, 2000; Volante, 2006), the use of an Achievement Test is generally questioned as a measure of predicting academic success. In view of the many inconsistent findings of predictive validity and admission standards studies, these issues continue to be of interest to researchers, and they need to be further investigated whenever it is possible (Linn, 1994; Alnahdi, 2015).

The current research confirmed this statement with findings showing the weakness of the predictive validity of the admissions criteria used in the College of Education at King Faisal University in general and the criteria of the QIYAS Centre tests in particular.

In order to achieve the aims of *Vision 2030*, universities in Saudi Arabia might consider reviewing their strategic objectives to marry up with this new phase of change in Saudi Arabia. This should include reference to *Vision 2030* in their strategic planning and mission statements in order to align with the broad objectives of the *Vision* and ensure long-term and sustainable quality in Higher Education. If those in Higher Education want to offer new skills to students such as creative skills, and learner-led and independent thinking (that are explicit within *Vision 2030*), then university curricula represent a place where that can be implemented and where an assessment of these qualities and characteristics needs to be considered.

Traditional educational methods are no longer adequate for teachers to be able to succeed in a highly competitive world (Quansah, 2018). The Saudi approach to education and learning is largely teacher-based, with teachers in charge of knowledge transfer and students preparing individually for traditional exams (Al-Hajili, 2016; Algarfi, 2010). The Saudi Higher Education system should also aim to overcome shortcomings in curricula and teaching methods and improve teacher training programmes and assessment methods. The current system refers to only four types of teaching methods that must be followed by teachers, namely lectures, discussions, dialogues and experiments (Algarfi, 2010; Sofi, 2015). According to Soanah et al. (2003), this inflexibility may discourage learners from being creative thinkers and can foster a lack of self-confidence and fear of criticism. This clearly suggests a need for a move to increase the quality of teaching, as well moving towards more student-led rather than teacher-led teaching.

7.6   Research question six:

In this section, the findings of the sixth question in this study are discussed. This question was: 'Is there a statistically significant relationship between the score of the General

Aptitude Test, the Achievement Test, and High School percentage with respect to the variables of gender and change of major?'

The results indicated that female learners exhibited better performance compared to males in both the overall High School percentage (girls mean = 96.00% while boys had 85.15%) and the General Aptitude test (girls' mean=77.81%, boys' mean=72.33%) in the Education College at KFU, and that the differences are statistically significant ((t(564.25)=-29.97, p<0.001) and (t(691.00)=-10.00, p<0.001) respectively).

In general, the average of predicting future academic success in exams taken and the GPA at the college level was higher for female Saudi students than for the male students, and a number of studies have also demonstrated such an outcome (e.g. Kim et al. 2003; Morgan, 1990; Young & Kobrin, 2001; Zwick, 2013). In fact, Morgan (1990) examined the predictive validity of the SAT, the Test of Standard Written English (TSWE), and Achievement Tests, which are prepared by the US College Board within groups classified by gender, race and college in 1978, 1981, and 1985. His sample study comprised 278,074 students from 198 colleges in the U.S. According to the results, females have higher correlation estimates than males.

Young and Kobrin (2001) reported on and analysed the available studies of differential prediction of admission tests over a period of more than 25 years (1974-2001). None of these studies were carried out in the context of participants who were completing their degree in an Education department. Consistent outcomes of differences in prediction among groups was found. For studies based on gender, these authors revealed significant differences between male and female students. Female achievement usually exhibited a higher correlation than that of males. More recently, Shaw et al. (2012) investigated the validity of the SAT for predicting cumulative Grade Point Average by college major. The authors examined more than ten majors, including education. However, the work of Shaw et al. did

not focus on the Education College as the sole discipline, and this study was based only on quantitative methods. The results indicated that the multiple correlations between the SAT and College GPA by major tended to be stronger for females compared to male students, with the main differences specifically found in education (female r = 0.62; male r = 0.52) Similarly, for most majors, there was a strong relationship between HSGPA and CGPA for female learners as compared to male students.

The picture is not consistent, however, as there are some studies that have found that male learners exhibit higher performance in predicting future academic success than female students do (Sireci & Talento-Miller, 2006; Wainer & Steinberg, 1992). The literature review on gender differences in academic performance at different levels thus reveals varied outcomes. Specifically, in Higher Education, studies show that female students are often found to outperform their male counterparts (Dayıoglu & Turut-Asik, 2004). In attempting to offer potential reasons for these differences in prediction, Leonard and Jiang (1999) and Weis et al. (2013) argued that female students have tended to have better study skills than their male counterparts. Both Wainer and Steinberg (1992) and Spinath et al. (2014) added that females received higher grades than males because they worked harder and attended lessons and classes more frequently. According to some university websites in Gulf countries, a pattern emerges in that the number of female applicants that are accepted onto courses within Higher Education is higher than for male applicants. For example, according to the Gulf Newspaper, just under two thirds of students who were accepted in the 2014 Academic year into the Emirati Higher Education system were female. Furthermore, at the Kuwait University, 37,234 students were accepted in 2014, of which 71% were female (The Office of the Vice President for Planning, 2014). According to the Education Ministry in Saudi Arabia, there were 866,823 students accepted in all colleges and universities in Saudi Arabia in 2010 and 477,370 of them were female. In addition, according to the KFU website, the university accepted 19,754 students in 2010 for bachelor and diploma degrees, and

12,050 of these were female. The study reported in this thesis showed that the population of female learners is higher than that of male learners in the Gulf countries and also at all the universities in Saudi Arabia. This clearly evidences that the Gulf and Saudi societies are particularly open to women's education, but women are viewed from a different perspective in the field of work (Varshney, 2019). It is possible that this view is related to the place of work where there is mixing between men and women or the type of work that is thought, stereotypically, not to suit the nature of women.

On the other hand, in the 2015 statistics in Saudi Arabia, 68% of unemployed Saudi females had received a bachelor's degree or higher compared to 21% of unemployed men. This is further evidence of women's educational superiority, but it appears that there is equally a corresponding failure to employ these women in jobs in keeping with their abilities (Koyame-Marsh, 2017).

The current research, which highlights the superiority of women when compared to men in academic attainment, both at the High School and university level, serves the best interests of women and the creation of jobs suited to their abilities in order to reduce the number of unemployed women, as well as investing in a very large body of skilled and educated individuals that has not yet been exploited in the building and development of the kingdom. This represents another clear contribution of this thesis.

For the second part of research question six, it is evident that those students who had changed their major had a higher High School percentage mean (94%) compared to those that did not change major (91%), and the mean difference was statistically significant with a p-value of 0.001. However, for the GAT and ACT, those that did not change their major had higher mean scores (GAT = 75.7 % and ACT = 81.4%) compared to those that changed major (71.2% and 69.9% respectively) and this mean difference was statistically significant with a p-value of 0.001.

This finding could add further weight to the fact that the High School percentage has the highest rate for expectation of success in the College of Education. Students may receive a high percentage from High School, but, due to the lack of high scores in the admission tests, they may not be accepted into the academic discipline that they wish to study. Therefore, after taking great efforts for the first academic term, they seek to change their major to one that they both prefer and which more consistently corresponds to their achievement in High School. As such, one could say that there is an effect of the High School certificate on changing subject. This may also be indicative of the need for the capacity test to be further examined and reviewed by the QIYAS Centre, and this will be addressed in more detail in the final chapter.

In terms of the variable of major change, this was introduced to determine its relevance to the admission criteria in the Faculty of Education at King Faisal University. The system of university admission in the Kingdom of Saudi Arabia is based on the calculation of a weighted percentage, including High School tests, the General Aptitude Test and the Achievement Test. Some students change their major during the first year of study, and this category makes up a significant percentage. As such, the researcher wanted to reveal the effect of this variable in the admission criteria used in the Faculty of Education. The reason for this is because a change of major could have either a positive or negative effect on the students' achievement. The researcher did not come across any studies carried out either in Saudi Arabia or internationally in relation to this subject when studying the predictive validity of the admission standards in general. It is evident from the responses in this section for Question Six that there is a positive effect for the students who changed their major in the first year of Education in the Faculty of Education on the overall percentage of High School. On the other hand, there is a reverse effect for major change on the increase in overall Aptitude Test Scores. There will be a more in-depth analysis of the reasons for changing

major and its impact on the students in the answer to Question Eight when addressing the views of faculty members in Education.

As the setting for this thesis is Saudi Arabia, the outcomes from this study are particularly timely because standardised tests are currently under review throughout the Kingdom and are being subjected to some 'experimentation'. Such tests have only very recently been introduced to the teaching sector and members of the public. Therefore, given these tests are currently required in all schools across Saudi Arabia, this study may provide a window of opportunity for those working on the preparation and design of these standardised tests to review the internal validity of these instruments and determine if they are currently fit-for-purpose (Al-Hattami, 2012). This study may also prompt some reconsideration of the admission criteria presently used by the College of Education to meet the requirements of *Vision 2030* which have been made explicit in the National Transition Document (2017). Of the many important items which are set out within Saudi's *Vision 2030*, a number are associated with Higher Education. These items can best be summed up as having a focus on improving teaching methods that position the learner at the heart of the learning process rather than learning merely being directed by the teacher. Such an emphasis points, therefore, to the need to build personal skills and encourage creativity. Saudi's *Vision* also focuses on the development of alternative assessment methods and curricula. This study is one of the first with implications for the delivery of *Vision 2030* in the Saudi Higher Education sector, as it has, for the first time, involved both genders in an Arabic context. This comparison is consistent with *Vision 2030* in terms of empowering women in the various fields of life/study, such as taking part in scientific research and, more explicitly, highlighting their success and academic achievements alongside those of men.

7.7   Research question seven:

As set out within the methodology chapter, this thesis, in addition to a quantitative dimension, included a qualitative element. Therefore, in this section, the chapter discusses the findings of the seventh question in this study which is: 'What are the College of Education Lecturers' perceptions about the current admission standards used at the College of Education at King Faisal University (The accumulative rate in the High School, the marks of the General Aptitude Tests and the marks of the Achievement Test)?'. Research questions seven and eight were answered using data collected using semi-structured interviews. The researcher interviewed eight academics who were currently teaching a number of Education courses to different academic levels.

Faculty members at the Faculty of Education at King Faisal University (regarding the admission standards currently used in the College when admitting students) reported that such standards are, in their opinion, insufficient and cannot alone be used as a benchmark in the admission process. One reason for being somewhat sceptical of these standards was the lack of studies and evidence that support the National Centre for Measurement and Evaluation (QIYAS) tests, with most of the respondents mentioning this specific reason. In response, some of the respondents went as far as calling for a review of the tests of the Centre of Measurement in terms of their validity and how the standards are explained.

While there was some hesitancy in being specific about which 'test', the more general concerns expressed by some lecturers about the reliability of tests developed by the QIYAS centre concur with the work of Al Hajri (2013, p. 308) which concluded that "most teachers and students did not consider that standardised tests were more reliable or valid".

Given that all QIYAS questions are multiple choice, some of the respondents made mention of some negativism that might affect the validity of the General Aptitude Tests and the Achievement Test, which, in their view, centred on the high percentage of guesswork.

The selected answer choices range from three to four, which means that the percentage of guessing the correct answer is between 25% and 33.3%. (Lesage, Valcke, & Sabbe, 2013). This percentage is considered high, which indicates that one-quarter or one-third of the final mark could very well be achieved by an individual student through guesswork. As such, the validity of these criteria in terms of admission can be affected.

Despite all the criticism, there seemed to be almost an admission that the criteria set for the Aptitude Test and the Achievement Test were 'here to stay', which can be explained by the fact that the tests adopted by the QIYAS centre remained the only available standardised tests in Saudi Arabia used for admission criteria. However, the argument that the Measurement Centre tests are the only ones available for application in Saudi Arabia does not mean they are necessarily fair and valid for academic success as a general acceptance criterion. Indeed the issue of standardised assessment has received some very recent attention that points to some changes in admission criteria. As the Washington website indicated:

> "The University of Washington has decided to eliminate standardised testing as a requirement for incoming students. Similarly, "admissions officers have studied results and results for several years and found standardised test scores have little correlation with success at the UW, particularly for Washington residents".

Similarly, some other institutions in the US have also taken the initiative to make this change alongside the UW. For example, it was decided by the University of California system in May to leave out the testing prerequisite up until 2024, thus allowing the school management time to come up with a fresh test or reject the prerequisite once and for all (Therrien, 2020).

While there was no reference to removing standardised testing in Saudi among the participants who were interviewed, interestingly, some of them in this thesis put forward additional criteria to render the process of admission to the Faculty of Education to be, in

their opinion, fairer. In effect, the use of an interview would be an addition and not a replacement. Specifically, among these criteria was the face-to-face interview for students who were applying to study at the Faculty of Education. The need to add an interview tool to the admission criteria of the Faculty of Education at King Faisal University is, in fact, supported by the literature. For example, Edwards et al. (1990) considered the interview be one key criterion for admission to Higher Education because of its potential to measure skills and knowledge that cannot be measured by other criteria, such as fluency in speech, communication and the retrieval of evidence and experience by applicants. Furthermore, Lamadrid-Figueroa et al. (2012) argued that personal interviews can be used effectively as criteria for undergraduate admission as well as the score for the High School certificate. More recently, Oranye (2016) also found that the use of the interview tool in student selection seemed fair and objective in the professional health programmes at the University of Manitoba in Canada.

Further admission criteria were suggested by teaching staff in the Faculty of Education. A test, designed by a committee formed from each department in the Faculty of Education under the supervision of the Dean of the College, seemed desirable. Beyond a test that would measure students' knowledge and skills in the disciplines of the Faculty of Education, little further detail was offered by the lecturers. However, there was a view that this 'test' might also measure student familiarity with modern assessment methods.

As for the most appropriate admission criteria currently used in the Faculty of Education, the majority of respondents agreed that the High School test remained the best predictor of success in the Faculty of Education. Their view was consistent with the quantitative analysis undertaken to answer Question 4 and Question 5 in this thesis. This finding is also in keeping with the majority of previous studies employing quantitative approaches which, collectively, have verified the predictive validity for university admission criteria (Albanai, Balhadi &

Alkhawli, 2004; Almokhlafi 2001; Almiqdadi, 2010; Alshumrani, 2007; Alshehri, 2011; Darwaza 1987; Murshid, 2013).

According to the Ministry of Education website, the High School score in Saudi Arabia is calculated for the last two years of High School education (Year 2 and Year 3 of High School). This is often more comprehensive because it is the outcome of two years of study rather than a single test that is offered once or twice to the learner and upon which he/she is judged. Interestingly, one of the interviewees insisted on reviewing the percentages calculated for each of the admission criteria in the Faculty of Education. Specifically, they pointed out that the percentage of the General Capacity Test and the Achievement Test in the Faculty of Education should be reduced instead of reducing the High School percentage. An assistant who had studied overseas said that, in his interviews with some parents, they were told that the percentage allocated to the High School scores in the admission criterion for the Faculty of Education is low and should be higher than 50%.

7.8   Research question eight:

In this section, the chapter discusses the findings from the eighth question in this study which is: 'What opinions do the College of Education lecturers have about the relationship between assessments undertaken at the end of High School and those within the College of Education?' Analysing the interviews with the participants led to the identification of four key themes. As a reminder these were (1) Traditional methods continue, (2) Lack of variety of assessment, (3) The relationship between High School assessment and Education College assessment already exists, and (4) Knowledge is priority.

All participants pointed to the fact that traditional assessment methods were still used in most classrooms within the College of Education. According to the suggestions made by the participants, there was no difference claimed between the assessment methods employed at the High School level and those used within the College of Education at KFU. This result

does, however, call into question the difference (if any at all) between High School assessment and Higher Education assessment in terms of the approaches used to assess learners. The work of Alnahdi (2015) offers some comment on this. The author states: "…teaching at the university level is an extension of the secondary stage in terms of teaching methods and the expectations of students and focuses on the same levels of goals that are highlighted in the secondary stage" (p. 5). The majority of faculty members interviewed from the Education College diminished the importance of the tests from the QIYAS Centre in predicting the success of students in the Faculty of Education because of the lack of sufficient evidence of studies and research to support such views. Participants in this study were consistent in their views on the current patterns of assessment tools used.

If there is no clear and positive difference in favour of university education, it means that the difference between High School and the Faculty of Education is only in name and not in the assessment tools and content. Among the current policies of Saudi Higher Education is to diversify the methods of admission and tools of assessment used in order for such assessment to be genuine and fair. However, beyond some apparent but informal and non-authorised use of alternative assessment by some lecturers in the College, it would seem the range of assessment used was far from diverse. *Vision 2030* emphasised the importance of developing the abilities of the Saudi people and equipping them with qualitative skills and knowledge. Such *Vision* considers these skills, in particular, among the most important resources of the country, by stating that: "The skills and capabilities of our students are among our most important and valuable assets" (*Vision* 2030 in Education, 2020). Perhaps universities and Colleges of Education in particular are responsible for achieving this goal by re-reviewing the methods used and learners' assessment tools which, as shown in the results of this thesis, are still traditional from the viewpoint of the faculty members.

As stated in Chapter 2, in relation to current education policy in the Kingdom of Saudi Arabia, the educational process is based on several factors, including curricula, teachers, and

teaching methods, and the examination of the test results using various evaluation methods to ensure the credibility of the tools (Ministry of Education, 2007). The diversification and innovation of student assessment methods is outlined in Saudi's *Vision 2030*. As made explicit on the official website of the Ministry of Education, one of the means of change that this *Vision* seeks to achieve in Higher Education is the development of systems and procedures aimed at ensuring rigour and discipline in the educational system, as well as promoting fairness in the assessment process. Some questions over the reliability of some current standardised assessments at KFU were outlined by several lecturers but, at this point, there seemed to be limited attention to a broader range of assessments within the College that might be seen as potentially 'fairer'. Processes that seek the 'voice' of the student as a contributor to this position are currently lacking. It was not possible to interview students as they had already left the university (after the relevant database period), but it would be interesting to interview current education undergraduates to gather their views on 'fairness' and how they were assessed at High School and how this form of assessment and what it was designed to measure transferred to learning and assessment experiences in Higher Education.

One of the reasons cited by participants in the interviews for a continuance in the use of these traditional methods of assessment was the large number of students in any one classroom, as well as the perceived lack of appropriate ability of students whenever more contemporary methods are used in assessment. As has been observed elsewhere, very large numbers of students in a classroom has posed a problem for the educational system in general and a challenge for its teachers in the assessment and evaluation procedures that can realistically operate (Iipinge, 2013). Of some relevance to this thesis, Bati et al. (2014) noted that one of the most significant problems facing the process of evaluation in Higher Education in developing countries is the large number of students in the same classroom. This may be due to the foundation stages of education in Saudi Arabia which are focused on traditional

methods of assessment which tend not to take into account individual differences and preferences among students and are often based on memorisation and skills of retention (Al-Harthy, 2010). A further reason cited by some participants in the interviews for a continuance in the use of conventional assessment methods was that (some) students could possibly be uneasy about being assessed by what, for them, could be regarded as 'new' methods. Some lecturers even questioned whether these learners had sufficient 'skills' or the necessary 'work ethic' to be assessed by new assessment methods, as was mentioned in the previous chapter. One of the teaching assistants stated that the reason students did not want alternative methods of assessment was that they did not have sufficient skills for such methods. This may be due to the fact that they have not been used to such methods of assessment since an early academic stage.

Although interviews were not employed with undergraduate students in this thesis, something about their response to new assessment tools can be inferred from the small number of accounts from some lecturers. This was the case in one participant's reference to losing almost a third of the class very early in a module where a presentation was included as a mode of assessment. It is fair to say that these forms of assessment (e.g., presentation and group work) are new to students as they have quite likely become used to being assessed through exams and multiple-choice tests and the fear of alternative assessment might be one of the reasons for their choice to withdraw from a given module. Therefore, based on the answers of the lecturers, it is quite worrying that some students might decide to take one class over another due to the type of method used by lecturers in their assessment. It is extremely important that these learners to come to accept these new assessment arrangements, and it is vital, therefore, to expand and deepen the university teachers' knowledge base regarding alternative assessment methods (Janisch, Liu & Akrofi, 2007).

Individual attempts by some to use more modern assessment methods such as WhatsApp and peer assessment were infrequent. That these forms of assessment had not yet been

formally agreed by the administration of the College of Education was evident in some responses and an expectation of compliance was mostly felt, but perhaps less so by some. Lecturers who studied in Saudi Arabia were more inclined to support traditional forms of assessment, whereas lecturers who had studied outside Saudi Arabia tended to experiment with the alternative assessment methods in an informal way because they believed in their importance in developing the students in the Faculty of Education. Of relevance, as an assessment and learning tool from the viewpoint of learners, WhatsApp has been reported by some as a modern and easy-to-use medium the world over (Gon & Rawekar, 2017). Also further noted by these authors, WhatsApp has the potential to support increased focus on student creativity, independence, and responsibility for learning. Marshall (2002) adds that techniques that mix images, texts and sound within an on-line platform can simultaneously help information retention by more than 50%. The outcomes from the few efforts reported at KFU were not forthcoming and the relevant participants in this thesis claimed they were both wary of students' possible reactions and the comments and reactions, particularly if unsupportive, from their teaching colleagues.

In view of the top-down structures that exist in the case institution and in Saudi Arabia more generally as a centralised and conservative system, it was perhaps unsurprising that the majority of respondents appeared to accept that the assessment methods currently used in the College of Education at KFU were aligned with the mandatory assessment methods used in High School. Nevertheless, suggestions for change did emerge, and initial teacher education was seen as a possible location for this with alternative assessment included as part of the programme. Teachers in general education could attend one or two academic terms in one of the Colleges of Education in Saudi Arabia. Here, the College of Education would include attention to 'other' methods of assessment and meet the expectations of the Saudi Arabian *Vision 2030*. Participants in this thesis also saw a role for the Ministry of Education in the development of programmes and training courses that are compulsory for

teachers and educational leaders at all education stages, especially focussing on assessment methods (Alzahrani, 2010). Such courses might specifically focus on understanding alternative assessment methods, their advantages over other traditional methods of assessment, and instructional strategies to implement them. This input may encourage the targeted teachers on these courses to adopt them and then use them in the classroom with the support of their senior management. This procedure may have a positive impact on the foundation and formation of students in public education and their acceptance of alternative assessment methods.

Quite a strong criticism expressed by the junior lecturers towards the traditional methods of assessment was that, in their opinion, they did not measure higher areas of thinking, but, instead, focussed on the minimum level of knowledge and skills. One of the participants cited a study where he concluded that 80% of faculty members' questions measured memory and 20% measured the level of application. Some lack of support for the current curriculum in terms of measuring higher skills emerged, and the diversification of methods and tools of assessment at KFU was seen as possibly contributing to the measuring of higher levels of learners' knowledge and skills. Perhaps what is mentioned in the literature review chapter (Chapter Three), especially with regards to the importance and benefits of alternative evaluation, is appropriate in this field in order to increase the awareness of faculty members in the College of Education about it so that they are able to apply it in the near future.

Of note is that these criticisms of traditional tests have been referred to by other researchers. Eckes (2007), for example, reported that the majority of traditional methods of assessment measure students' basic cognitive skills. Smaldino et al. (2000) similarly noted that traditional assessment methods measure the students' ability to memorise. This type of assessment is not sufficient to assess higher cognitive skills, such as problem-solving and critical thinking (Birgin & Baki, 2007), which *Vision 2030* ironically emphasises. Where a wide range of assessment tools are employed, students' skills, such as problem-solving,

critical thinking, and analysis skills, can be gauged, as well as being able to determine if students can apply their knowledge to solve new problems and express themselves well verbally and in writing (e.g. Shepard, 2000). There is a great importance attached through *Vision 2030* to innovation and entrepreneurship, as well as to the quality of the citizens that Saudi Arabia wants to produce in years to come. In particular, this relates to citizens having the ability to engage in problem solving and collaborative teamwork, and to develop self-confidence traits. Higher Education is seen as a major setting to develop skills and abilities in high quality graduates which are compatible with the labour market in Saudi and which fulfil some key objectives of *Vision 2030.* From the data, few efforts to change the traditional assessment used in the Faculty of Education to evaluate such 'learning' was evident, and, where mentioned, this has not been noticeable. How decision-makers in the Faculty of Education would respond to this change of method in assessing students is a concern.

Drawing on the analysis of the responses provided by the faculty members, there are two further themes worthy of discussion. The two themes are (1) the change of major or subject and, (2) the educational achievement of women.

Question 6 found a statistically significant relationship between students who changed their major and the High School average. At the same time, some members of the Faculty of Education mentioned that the impact of changing the subject depends on the personality of the learner, which may be positive at times and negative at others. Thus, the whole process of changing subjects is greatly dependent on the personality of the learner and their own personal wishes and circumstances.

It would seem that the current system where students are allocated a major depending upon their High School achievement score does not always suit some learners and it was not surprising to learn from the data that when the student eventually followed their preferred major, their future educational performance was better. Saudi might consider reviewing the

present arrangements so that students do not end up in a major that, according to their High School scores, they are suited to academically and can cope with, but is one they neither wish to follow nor enjoy and, thus, end up changing majors. This matter is about the ability to choose and introducing some flexibility within the admission system. We need to educate students at the secondary level on the horizons of the job market and the appropriate disciplines so that learners do not have to change majors during university study and thus lose time (one year and sometimes longer) to graduate.

Saudi Arabia is anxious to improve its standings in the world university league tables. Having knowledge of the relationship between High School performance and university performance for individual boys and girls is necessary to inform any coming changes in assessment/policy. *Vision 2030* is a key step change for Saudi education and, in particular, for women. The second theme that emerged through the analysis of the qualitative data was the high academic achievement of female students in comparison to their male counterparts. All participants tended to state that girls often performed better in their academic achievement than boys but the reasons for this varied from one participant to another. While some stated that customs and traditions may be a primary cause, it was evident that some quite traditional views of females and their role in Saudi Arabia were held, such as females spending most of their time at home. This was offered by some as a reason why girls do well in school. Others noted that job opportunities for girls in Saudi Arabia have traditionally been few and far between, and this was put forward as one reason why female students worked hard in a fiercely competitive environment for higher-level academic qualifications in order to gain access to the job market, as was stated in the findings chapter. Thus, we can say that the employment opportunities for girls are limited because of the cultural aspect of Saudi society according to some of the participants' perceptions.

Saudi opportunities for girls/women both socially and culturally are changing, so some of the comments about women staying at home could be seen as dated and very likely not

applicable in the future. More than one participant mentioned that a further reason for the high achievement of female students was the level of academic competitiveness among women compared to male students, which, in their view, could be a factor explaining the performance of women.

Some of the above reasons have been reflected in the results of studies conducted in Europe and America, which also concluded that female students were academically higher achievers compared to their male counterparts. In general, the average of predicting future academic success from exams taken and GPA at the colleges were higher for female students than male students (Kim et al. 2003; Morgan, 1990; Young & Kobrin, 2001; Zwick, 2013). In a study by Voyer and Voyer (2014), titled Gender Differences in Scholastic Achievement, researchers from the University of New Brunswick relied on a meta-analysis of 369 studies conducted between 1914 and 2011. These studies included the academic grades of over one million boys and girls from more than thirty countries (including the United States, Britain, Germany, Turkey, India, Australia, etc.). All studies included an assessment of gender differences in teachers' grades or GPAs in general, and undergraduate or postgraduate education. The study found that girls' grades were consistently higher than boys' over several decades without any major changes in recent years. As to why girls performed better in school than boys, the authors speculated that social and cultural factors could be among several possible explanations, which, from the lecturer's point of view in this thesis, may have been a reason. Parents may assume that boys are better at maths and science, so they may encourage girls to do more in their studies, which may be an advantage for girls across all courses. Voyer and Voyer (2014) also showed that girls tended to study in order to understand materials, while boys emphasised performance or outcome, indicating concentration on final scores. According to these authors, "mastery of the subject generally results in better marks than focusing on performance, so this can be partly attributed to the males having lower scores than females (p. 1192). Seligman and Duckworth (2006) found

that middle school aged girls outperformed boys in self-discipline, which may contribute significantly to better grades in subjects. They also found that girls were more proficient in reading test instructions before starting to answer, preferred homework to watching television, and continued with longer tasks despite boredom.

In contrast, there are some studies that found males to have higher performance in predicting future academic success than females (Sireci & Talento-Miller, 2006; Wainer & Steinberg, 1992). The literature review on gender differences in academic performance at different levels therefore reveals varied outcomes. Higher Education studies show that females often outperform males (Dayıoglu & Turut-Asik, 2004). Leonard and Jiang (1999) and Weis et al. (2013) argued that female students have better study skills than their male counterparts. Wainer and Steinberg (1992) and Spinath et al. (2014) added that females receive higher grades than males because they work harder (which may be implied by the Saudi lecturers in this thesis) and attend lessons and classes more frequently. Of late, some researchers have stated that a great deal of research remains to be done to identify gender-related factors in school performance, as well as their possible causes (Unity & Igbudu, 2015).

Regarding *Vision 2030,* we have seen some of its effects by empowering Saudi women in some fields, for example, we have more than one ambassador for the Kingdom of Saudi Arabia. Also, the Cultural Attaché in Britain for the first time in Saudi's history is a woman. This appointment is an indicator of the quality of Saudi women in the political and cultural spheres, whenever opportunities arise for them.

The next chapter presents the conclusions of the research and also includes some recommendations from the findings, possible areas for further research and also acknowledges some limitations of this thesis.

Chapter 8 : Conclusion

The previous three chapters presented the findings and then discussed these in relation to relevant literature. This chapter outlines the major conclusions which can be drawn from the study. The chapter then summarises the original contributions of this research. This is followed by a presentation and discussion of some of the key implications for the study and, at the same time, also highlights some limitations of the research. Suggestions for further research and development are then put forward.

8.1   Main findings

The main findings in this thesis indicate a statistically significant relationship between the three admission standards that are used at the Education College in KFU (which are the accumulative rate at High School (SGPA), the marks from the General Aptitude Tests (GAT) and the marks from the Achievement Test (AT)) and the accumulative rate from the College of Education (UGPA). However, the strength of each relationship varies from one criterion to another with respect to the UGPA. To illustrate, there is a statistically significant relationship between SGPA and UGPA at KFU ($r=0.562$, $p<0.01$), between the GAT and UGPA ($r=0.32$, $p<0.01$) and between the AT and the UGPA ($r=0.27$, $p<0.01$). However, the findings indicated that only SGPA and GAT significantly predicted the GPA for Education ($p<0.0001$) but the AT scores are not significantly associated with UGPA if these are estimated together with the SGPA and GAT ($p>0.05$). This means that the SGPA and the GAT are the most influential admission standards in predicting university GPA within the Education department at issue in this thesis.

The results also indicated that the SGPA is the most important factor in predicting the future performance of students in the Education College (Beta coefficient 0.512), followed by GAT (0.163) and then AT (0.001). These results show that using High School GPA in conjunction with the Aptitude Test could help admissions employees to make more accurate predictions and more appropriate admission decisions than using High School scores alone.

According to Al-Ahmadi (2011), "[t]he government of Saudi Arabia has adopted a clear vision for the empowerment of women as reflected in recent development plans that show a clear shift in the orientation of planning efforts towards the development of women's roles instead of focusing on women's right to education and employment" (p.149). Although greater numbers of women are entering Saudi Higher Education than men, empirical research on women in Higher Education in Saudi Arabia remains very scarce indeed (Alomair, 2018; Khan, 2019). Long standing traditions sand cultural barriers in society could well be one of the reasons. However, *Vision 2030* may set a new path for empowering women to contribute to both the educational and economic growth of Saudi Arabia (Naseem & Dhruva, 2017). These authors also report that "Saudi women represent a wealth of untapped potential for the economy. Many of them are yet to join the labor force, despite being highly educated and motivated" (Naseem & Dhruva, 2017; p.23). Therefore, *Vision 2030* may contribute to changing the lifestyle of women from sitting at home, caring for children and limiting themselves to working in nursing and teaching to, in time, competing with men in other jobs, including engineering, law, architecture and other professional careers based on their academic achievement (Naseem & Dhruva, 2017).

This research has thus sought to shed light on the educational achievement of males and females in the Faculty of Education at King Faisal University. The results also indicated that females performed better compared to males in both the GAT and the overall SGPA in the Education College at KFU and that the differences are statistically significant. The reasons for this given by the male faculty respondents tended to refer to some taken-for-granted

assumptions about the role of women in Saudi society rather than the attitudes of women to academic study. In the main, the majority of participants argued that customs and traditions may be a key reason, with females in Saudi Arabia spending most of their time at home, and this may help them attain significant achievements. They also added that job opportunities for women in Saudi Arabia are far and few between, and this may explain why women work harder to compete for higher academic achievement in order to gain optimal access to the job market. However, the role of women in Saudi society is undergoing some changes in light of the *Vision 2030* of the new Crown Prince, so the future could see an expansion of opportunities for women which might also include wider access to learning, Higher Education and employment. Indeed, some have indicated that "[w]omen's empowerment seems to be topmost on the agenda of Saudi government" (Naseem & Dhruva, 2017, p.24).

As for any change of major, the results showed that those individuals who had changed their subject attained a higher mean SGPA score compared to those that did not change major, and the mean difference is statistically significant. However, for GAT, those that did not change major had higher mean scores compared to those that changed major and this mean difference is statistically significant. The current system where students are allocated to a major dependent upon their score at High School may require some revision as it is possible that some students may have achieved well in High School but end up enrolling in a major that they would prefer not to. It would seem that the system in place where students are allocated a major depending upon their High School achievement score, does not always suit some learners and it was not surprising from the data that, when the student eventually followed their preferred major, their future educational performance was better.

The last two research questions (seven and eight) were answered using the semi-structured interview technique and these questions concerned the perceptions among lecturers regarding assessment, predicting future performance, and the current position around assessment techniques in the College. Regarding the faculty members' perceptions at the

Education College in KFU about the admission standards currently used in the College, participants reported that such standards are insufficient and cannot be used as a sole benchmark in the admission process of students in the Faculty of Education. One reason for the insufficient nature of these standards is the lack of studies and evidence supporting the National Centre for Measurement and Evaluation's current regime for (QIYAS) testing, with most of the respondents mentioning this specific reason. This thesis offers some evidence along the above lines.

All participants pointed to the fact that traditional assessment methods were still being used in most classrooms within the College of Education at KFU. It is clear that the attitude of faculty members from Education at KFU is towards the existence of a link between the assessment tools from High School and those in the Faculty of Education through the use of the same traditional methods of assessment at both stages. The participants' view is also supported by previous criticisms of traditional assessment techniques in that they are not so important in terms of measuring higher levels of skills and knowledge (Birgin & Baki, 2007; Quansah, 2018).

One of the reasons cited by participants for the continuation of these more traditional methods of assessment was the large number of students in the classroom (Khan & Iqbal, 2012; Moghal et al., 2019). Participants also referred to a perceived lack of appropriate abilities/skills among students to cope with some modern methods that are used in assessment (oral communication/group assessments), which very likely did not feature in their education prior to beginning university study. If participants are generally happy with the annual performance of their students upon completion of their degree, they may possibly see no reason to change the ways in which they are assessed. This reluctance to change might also be attributed to an inability to influence substantial reform of the manner in which high school students are currently assessed. There was indication that participants who had

completed higher degrees overseas seemed more aware of alternative assessment tools than those educated entirely in Saudi Arabia.

Summarising the conclusions around the on-going use of traditional assessment modes in the College, there were some examples of some participants employing alternative assessment, but these were not formal efforts and perhaps were 'hidden from view' and did not align with the current policy within the College. These efforts tended to come from younger lecturers and those that had completed degrees outside of Saudi Arabia. However, these lecturers faced some challenges that prevented them from continuing to apply these types of alternative assessments. Some of the students withdrew from courses when the lecturers used alternative assessments; this might be because of a lack of prior experience among students with these forms of assessment. This is one of the potential challenges which may lie ahead - orientating these students to being assessed 'differently'.

If the Faculty of Education wants to make changes to assessment they might consider some pilot work, perhaps with a smaller class given that large classes seemed to be a barrier to trying other assessment methods (e.g. group work). It might be worth speculating here whether there could be a risk to the College in trying new forms of assessment. The students might perform less well when not doing exams or multiple-choice tests. Could this performance impact upon their satisfaction and GPA? There could be some resistance to change from some lecturers and from the College as a whole. Young lecturers in the College of Education should also organise workshops on some of the assessment methods in order to encourage the rest of the teaching staff to adopt them and to relay their advantages to the students so as not to be taken by surprise during the lectures. Given his job position and his predisposition towards alternative assessment methods, the Dean of the College of Education is required to assume a role in terms of asking the administration of King Faisal University to allow faculty members in the Faculty of Education to use alternative assessment methods

because of the benefits discussed in this research and their strong association with the achievement of some of the objectives of *Vision 2030*.

The university is supposed to deal with this situation by cooperating with schools in public education, which could be achieved by providing secondary school teachers with training courses on the types and methods of alternative assessment, and familiarising students in secondary schools with alternative forms of assessment that they may expect at the university level. It may be useful to mention that a preparatory year for students at King Faisal University could be an option whereby alternative assessment methods are used to measure academic achievement in order to prepare students and acquaint them with alternative assessment methods before embarking on the desired university majors.

8.2   Limitations of the study

The qualitative sample had to be of sufficient size to include all those who were able to speak authoritatively about issues of assessment and predictive validity within the case institution. The individuals who were interviewed held a range of roles within the Education department, represented a variety of academic ranks, and had differing years of experience lecturing in Higher Education. Participants either had or had not taken the QIYAS tests themselves at the end of High School. The number of participants in the current research study was quite modest but represented the majority of available staff with duties in the teaching of Education modules. A larger number of participants would have been more representative. The study results therefore may not generalise to other Education departments in Saudi Arabia because the research is directly related to just one Education department in KFU. Nevertheless, the outcomes of the thesis would very likely be of interest and might encourage other Education departments in Saudi to engage in a similar analysis of their students' entry and exit data. Moreover, the people who were interviewed were those available to volunteer,

and so the researcher cannot thus generalise these results to anyone other than the people who volunteered to be interviewed.

Quantitative data was collected from students' files within the Admission and Registration Office at KFU. A pre-existing database was accessed. This database contained academic information about individual students and was located within the College of Education. This pre-existing database included for each student: High School percentages, General Aptitude Test scores, Achievement Test scores and cumulative final year College Grade Point Averages. These data were extracted from the students' files (for both male and female students) for students who were admitted in 2010 and graduated in 2014 from the College of Education at KFU. The data did not include students who left the Education College at KFU before the end of the academic year 2014 and any students who had not graduated by the end of the academic year, 2014. In addition, the database retrieval did not reveal any personal information about individuals (names, ages etc.). It only revealed information about their assessment scores. No one was able to access the data except for the researcher and the research supervisor who had access to the data presented in the findings during the writing-up of the study. The data obtained from this study were not passed to any third parties.

8.3   Original Contributions of this thesis

Internationally, there are very few studies specifically addressing predictive validity among samples of university-level students who are studying for degrees in Education. As the literature review indicated, the vast majority of empirical studies lie in Management, Engineering and Business (Kass, Grandzol & Bommer, 2012; Shaw et al., 2012; Sireci & Talento-Miller, 2006; Wait & Gressel, 2009). This focus upon particular academic disciplines is similar to the position within Arab countries (e.g. Alqatei & Alharbi, 2012).

Based upon a review of the literature, it could be concluded that there were no predictive validity studies in the Arabic world which included both male and female Education students as participants. A study in a strongly conservative country such as Saudi that includes data from both men and women is unique. In Saudi Arabia, given the separation of men and women in education, an original contribution would be a study involving females as participants, which was indeed the case in this thesis. As a result, we have little, if any, understanding of the performance success of women in terms of predictive validity in Education based upon their academic achievement at the end of High School. This inclusion of data for comparative purposes on the basis of gender, it is argued, is seen as a part of the changing culture in Saudi Arabia around the widening of academic opportunities and achievement of females. The percentage of Saudi women in the workforce still remains quite low in Education (Rajkhan, 2014). Some time ago, Doumato (1999) attributed the reduced percentage of women in the workforce to segregation between the genders. Segregation might not be the only reason for this, however. For instance, there is a complete absence of studies that solely investigate the results of Saudi girls' achievements or the extent of success of women in criteria for prediction. We need to study Saudi women in more depth within Saudi society to determine needs for future development and to provide the necessary educational opportunities for women based on their abilities and academic successes. Thus, this study is, in part, an opportunity to understand what we know about the prediction of future success for women in this context.

Although the percentage of Saudi women in the workforce still remains quite low in Education (Rajkhan, 2014), it has been reported that "[w]ith over 50% of university graduates being female, Saudi government will continue to develop their talents, invest in their productive capabilities and enable them to strengthen their future and contribute to the development of society and economy" (Naseem & Dhruva, 2017, p.26).

It is argued that, if the College in this study has clear evidence of the performance of students based on their incoming assessment scores, then they will have more information to determine whether the nature of the assessments they are currently using within modules are appropriate. The participants were of the view that the types of assessment being used in the education modules were appropriate given the manner in which pupils were assessed at the end of High School. Predicting future performance based upon a combination of admission variables gives the KFU Education department valuable comparative evidence as to the relative strengths of prediction. To that end, the findings could potentially support decision-making about the possible development of assessment processes as appropriate. Studies employing a sequential mixed methods approach to examine predictive validity are also few.

This study may also help the Education College at KFU in terms of meeting its growing concerns about the ranking of the university both domestically and internationally on the basis of the performance of students. The interviews provided an opportunity to engage in conversations with academics regarding the prediction of student performance and this approach might be of interest to other GCC states in Education and also to other academic departments within the case university.

This research is very timely given the new *Vision* in Saudi Arabia termed "*Vision 2030*" and developed by CEAD* (2016). *Vision 2030* has been implemented since late 2016 to develop the Saudi country in many aspects, including health, business, technology, industry, tourism and education. One educational element of this *Vision* is the aim of having five Saudi universities in the top 200 universities in the world by 2030. This aim is likely to lead to strong competition among universities to attract the best students who have the potential to be successful graduates. The universities will therefore need to demonstrate strong graduation outcomes in order to contribute to "*Vision 2030*". It is hoped that having a clear picture of the relationship between High School performance and university performance will assist Saudi universities in achieving high quality recruitment that could lead to strong

academic outcomes which would then positively affect their standing in terms of the quality of the student experience (Buela-Casal et al., 2007; Khosrowjerdi & Kashani, 2013; Safón, 2013) and thus meet one aspiration of *Vision 2030*.

In addition, there still is some inconsistency in the research that has addressed the prediction of future academic performance. There is much debate, in the literature, about which of the standardised test scores - Grade Point Average of High School percentage or standardised examination - should be weighted more heavily when making admission decisions. A number of studies have argued that High School Grade Point Average more accurately predicts academic success in colleges than standardised tests (Alshehri, 2011; Alshumrani, 2007; Fleming & Garcia, 1998; Kobrin et al., 2008; Murshid, 2013; Snyder, Hackett, Stewart & Smith, 2003; Vulperhorst et al., 2018). However, some studies show that standardised tests, such as the SAT or QIYAS, more accurately predict college success than High School Grade Point Average does (Alqatei & Alharbi, 2012; Geiser & Studley, 2010; Morgan, 1990; Shaw et al. 2012). Given that 'Education' is not a subject in and of itself that is 'taught' in High School or might be a subject option for a pupil at the end of High School, the findings from this study offer a contribution to the knowledge about which predictor is more significant in predicting students' college/university success, but with a specific and unique focus upon Education undergraduate students.

Furthermore, it has been noted above that none of the previous similar studies in the broader context of this thesis have included a mixed-methods approach in an Education College, nor have they focussed on both genders in the investigation of the predictive validity of the admission standards, nor whether there would be a variation in terms of gender and changed major. It is therefore likely that the current study would be the only one to use mixed methods to address variation of the predictive validity of the standards in an Education College only in terms of gender difference.

8.4   Implications of the study

This section highlights the implications of the present research for assessment in High School, QIYAS central, and the Education College at KFU. It begins with implications for the structure and purpose of assessment in High School in general. Then, implications for QIYAS stakeholders are introduced. Finally, the section discusses the possible implications for the Education College at KFU.

The results of this study may assist High School teachers in reviewing the assessment methods they adopt when assessing student achievement. The results of this study demonstrated that the traditional methods of assessment using pen and paper are the main type of assessment employed in high schools. This type of assessment is mostly focused on measuring the minimum skills of knowledge as indicated by the majority of respondents to be very likely acquired through rote learning and memorisation. This study may also help High School teachers in terms of identifying other methods of assessment. Alternative assessment is one of the methods used in some developed countries, including discussion groups, presentations, and so on. These assessment techniques can help the learner gain confidence in public speaking, develop his/her skills in working cooperatively and solving problems, as well as measuring higher levels of thinking. This study may also contribute to the revision of the assessment objectives in High Schools and the implementation mechanisms thereof. Research indicates that assessment in High Schools tends to encourage learning for the purpose of sitting exams and not for achievement, knowledge and creativity, which, according to some researchers, is a key disadvantage of standardised tests in general (Smith & Fey, 2000; Volante, 2004, 2006). A common phrase - 'teach to the test' - may be used here. It is important, therefore, to rethink the school environment in High Schools in Saudi Arabia in general, which requires a review of the assessment goals in particular to make it more attractive and beneficial for learners both now and in the future. There are,

however, some challenges. Some respondents in the Faculty of Education at KFU complained that students were not familiar with alternative assessment methods and this point tended to deter the university participants from introducing them. In addition, the participants even claimed that their presence led to a number of students withdrawing from their modules. Even if some university lecturers were keen to include alternative forms of assessment, they were somewhat fearful of using these methods and that they would be 'found out', which has consequently led some teachers who were interviewed for this thesis to stop using alternative methods in the assessment of students.

Thus, it is important therefore to review the assessment methods used by teachers both in general and Higher Education. In order to change the current methods of evaluation in the Faculty of Education, it is necessary to train teachers on alternative assessment methods (presentation, student achievement portfolio, etc.) because the majority of students graduating from the Faculty of Education are qualified to teach in education. Teacher education programs in Saudi Arabia need to be developed in the light of modern trends and in line with *Vision 2030*, which aims to train and prepare teachers who are familiar with teaching methods, learner assessment approaches, and providing feedback (Alsalem, 2017). This training may contribute to getting learners used to these types of assessment methods and also to the non-leakage of students from the lessons, according to some of the lecturers in this study. Therefore, if these students do not receive sufficient training, skills and knowledge about alternative assessment types during their Higher Education stage, they will continue to use traditional tests.

This study was not able to determine whether the tests prepared by the Measurement Centre (QIYAS) should be changed, as these are standardised and required by Saudi law. Nevertheless, one of the impacts that this study might yield is that it can help the Measurement and Evaluation Centre in Saudi Arabia (QIYAS) review the credibility and suitability of the Centre's tests. Of interest to QIYAS, the results of this study have

demonstrated that the High School percentage is still the best predictor of success in the Faculty of Education. Many respondents (some of whom are members of the Measurement Centre) tended to doubt the reliability of the Measurement Centre test questions and so a discussion between Higher Education officials and those working in the High School assessment centre is recommended. The tests of the Centre are not new and, in fact, have been around for 16 years since their establishment, but these tests are still being used and are criticised for their validity, as mentioned by many respondents. Reviewing the tests and examining them further may be helpful as an outcome of this research.

Standardised assessment in Saudi Arabia is based on recall. However, the question that might be posed here is: are those skills that are assessed by standardised tests in line with what *Vision 2030* sees as the type of students the country wishes to produce? It is evident from the results of this thesis and other previous research that the tests of the Measurement and Assessment Centre in Saudi Arabia (QIYAS) reveal a weak link to academic success compared to the secondary school certificate. It seems time to review these tests and their relevance to the goals of *Vision 2030*, which seeks to prepare citizens with the ability to engage in communication and teamwork and are characterised by their higher thinking and critical thinking skills, as opposed to the current system according to which students in Saudi Arabia are assessed using traditional tests and measurement tests based on multiple choice which may not allow this to happen at this stage. Therefore, in order to close the gap between secondary education and university education, as well as between the outputs of Higher Education and the labour market, it may be appropriate to fully review the admission criteria used in the Faculty of Education, especially the tests of the Centre of Measurement (QIYAS).

What adjustments might be warranted to assessments at the end of High School and encountered during study for an Education degree? This is also a question for the Faculty members at the Faculty of Education at KFU, who should consider the potential for some variety in the assessment methods used to measure the achievement of students while at the

same time being confident that this will help students achieve well by the end of their Education degree. Not all students prefer written or traditional methods as there are those who are in favour of presentations and also those who prefer discussions and other modes of assessment. This thesis may have potential implications in terms of diversifying the assessment methods used by my colleagues at the College of Education. Many colleagues have stated that they are trying to use modern methods in assessment, such as WhatsApp and other alternative assessment approaches. However, these attempts need to be formalised further within the College of Education and will require formal approval from senior management. This illustrates the politics of Higher Education in Saudi Arabia and that the top-down characteristic of education in Saudi means that university staff have limited autonomy in how they assess students. It means that opportunities to make comment and suggest changes that might address issues regarding the suitability of High School assessments is limited at best. An important question to ask is this: Is the Education College assessing the character, qualities, soft skills and abilities that Saudi sees as necessary for the future generation, particularly given the aims of *Vision 2030* and also that just under half of current Saudi's are technically 'young'? Is reliance on multiple-choice testing and very traditional exam-based modes of assessment fit-for-purpose in view of this *Vision*?

If education departments were to adjust their curriculum and assessment, would they risk reducing the knowledge they have on performance prediction and perhaps have students who perform less well?

8.5   Recommendations:

In light of the findings, conclusions and limitations of this research, a number of recommendations can be taken into consideration when further research is carried out in this area.

There is a need to conduct a study to find out the views of the students who have passed the following tests: the High School assessment, the tests of the Measurement Centre, and the assessment methods in the Faculty of Education to identify the correlation between the multiple assessment methods and the differences between them from the point of view of students. It is evident that the student voice is absent in this thesis but given that the data were secured from students who were undergraduates up until 2014, the opportunity to interview a sample of these same students was not possible, as they had already graduated and left the university. To learn more about the issue of the relationship between High School and university assessment would require a separate sample, and these would need to be current students within Education. Future studies could interview High School students and university students to gather their perceptions on the suitability of how they are/were assessed at the respective stages to investigate whether current assessment systems at QIYAS are appropriate, in their view, in terms of assessing necessary skills and dispositions for future work.

The current study revealed the lack of correlation and link between the general aptitude test, as well as the achievement test, and academic success in the College of Education at King Faisal University. Among the recommendations is to conduct field studies and research in the field of marketing university services for the College of Education and measure the level of satisfaction of the beneficiaries of those services, which should hopefully result in identifying the weakness and improving the services provided by King Faisal University, whether in terms of admission standards, teaching methods, or other areas. Also worth

investing in is to conduct research on the predictive validity of the admission criteria used at King Faisal University on samples from other colleges such as medicine and engineering, etc. and comparing those results with the current study. Similarly, since the Kingdom of Saudi Arabia is seeking to achieve the goals of its *Vision* 2030, and one of its most important goals in higher education is to link university education to the labour market, it is important to conduct research on the predictive validity of the admission criteria at King Faisal University in light of academic success and the competency of performance in the labour market. Given that the current study has proven that the traditional evaluation is prevalent and that the alternative evaluation is still rarely used by faculty members in the College of Education, it is useful in this respect to carry out studies related to understanding the reality of the ownership of faculty members in the College of Education (university colleges) at King Faisal University and their practice of alternative evaluation within the university.

In order to improve university admission standards at King Faisal University and benefit from developed countries in that domain, it is also possible to conduct comparative research between the admission criteria used at King Faisal University and the admission criteria used at universities in other parts of the world who score highly in university league tables for undergraduate students and the relevance of those criteria to the labour market for both universities. As this study along with other studies have proven the superiority of women in academic achievement over males, it is important to open further horizon for women, whether in admission to university programs or in the jobs available at King Faisal University.

The researcher also recommends not only a direct replication of this study for students across a further three years within the database but also the undertaking of a study of the predictive validity and impact of different assessment methods in the Faculty of Education on student achievement. It is recommended that, each year, the Education College runs those statistics that appear each year from the database to look at the relationship between High School and

the end of year one, year two and year three in the Education College. The value of this might be linked to what the Education College can do for the assessment that is currently used in the Education College if it bears a weak relationship with High School. For instance, the Education College can change the assessment that has a weak relationship with the High School to improve matters. This may also generate value for the individual student when they feel there is an improvement, which may, in turn, improve student satisfaction, in line with the requirements of the *Vision*.

It should be pointed out in the concluding section of this study that the aims of this thesis were to attempt to identify the predictive validity of admission standards at the Faculty of Education of the King Faisal University in the Kingdom of Saudi Arabia. Perhaps the most important finding of the current study is the validity of the High School measure as the best criterion for predicting success at the Faculty of Education. Another significant finding is that the Aptitude Test criterion is characterised by a reasonable correlation in predicting academic success, while there is a somewhat weak relationship between the Achievement Test and GPA in the College of Education. Female learners are also characterised by high educational achievement compared to their male counterparts. This study also confirms that traditional assessment methods are still in place both in High Schools and in the Faculty of Education. At the end of this research, the Ministry of Education and the College of Education at King Faisal University are urged to review and diversify the methods of assessments used in both High Schools and in the Faculty of Education in keeping with the students' expectations and in line with Saudi Arabia's *Vision 2030*, which has as one of its key goals the improvement of the learning outputs of Higher Education to align with the outputs of developed countries in the world. The researcher also suggests that the QIYAS and Evaluation Centre should conduct a periodic review of the Centre's tests, particularly with regards to credibility, as such tests have become compulsory criteria for admission to the different colleges and universities in Saudi Arabia.

If the KFU is going to rank globally, they need to track the best students, the students who achieve well and students who are satisfied. Thus, they need to "get courses right" and the assessment should be appropriate. Moreover, there should be an appropriate connection between the courses and the labour market also. If the KFU wishes to contribute to *Vision 2030*, then some of the current research regarding how students are assessed, and the variety of assessment offered might be relevant here.

Appendix A   **Ethical Approved**

Submission Number: 25338

Submission Name: The predictive validity of the admission standards in the college of education at King Faisal University in Saudi Arabia

This is email is to let you know your submission was approved by the Ethics Committee.

You can begin your research unless you are still awaiting specific Health and Safety approval (e.g. for a Genetic or Biological Materials Risk Assessment)

Comments

1.Thank you for completing the necessary Ethics Application Form for Secondary Data Analysis. Given you are traveling on 1 April, please ensure your Consent Form is consistent with your Ethics Form and Info Sheet in stating that withdrawal can be at any point during the interview (at the moment, your Consent Form does not say that). Good luck with your study.

Appendix B   **Interview protocol**

The research title: The predictive validity of the admission standards in the college of education at King Faisal University in Saudi Arabia.

Date: **Place:**   **Interviewer: Waleed Alshammari**

Interviewee: **Position of interviewee:**

Briefly describe the study:

The purpose of the study is to investigate the predictive validity of the current admission standards applied at the College of Education at King Faisal University (KFU) in Saudi Arabia and explore which score among the current criteria used has the strongest contribution to students' academic success. Further, since this study attempts to include students from both gender groups, and very few studies have included both genders at the general level and to the best of my knowledge non has been done in the Saudi context. Before in the investigation, this study aims to explore any possible variation between the criteria items in terms of gender group. Also, since none of the previous studies have addressed the issue if students changing their major after being admitted to certain major initially at the university, this study attempt to explore the academic performance of students who changed their major after starting their university study.

Interview questions:

1/ How long have you been in Higher Education?

2/ What modules in the different year groups do you teach?

3/ Can you describe the different types of assessment that we use in the College of Education?

4/ What type of assessment do you use in your classes?

5/ What is the reasons of using this type of assessment? Could you explain this for me?

6/ What do you know about the types of assessment that use in the end of High School? Tell me as much as you can?

7/ What skills / knowledge do these assess?

8/ Can you give me some ways in which we are trying connect (align) the assessment method of the High School with the assessment within the college?

9/ In your view, which type of admission standards that we use in the Education College has a higher relationship with the Grade Point Average (GPA) in the College of Education?

10/ Have you been able to predict the performance of student of knowledge of their QIYAS score?

11/ In your opinion, is that an accurate prediction? Can you give me some examples?

12/ Do you think that, students who change their subject in the first year of their study may affect on their GPA? Can you talk more of that?

13/ As you know, QIYAS used multiple-choice but in Education College we use different types of assessment. In your opinion, what should we doing in the College of Education to help student to improve their ability to do other types of assessment?

14/ What skills are tested through the assessment we use in Education?

15/ What knowledge is tested though the assessment we use in Education?

16/ What links are there between assessment at the end of High School and those in the College of Education?

17/ Do you think that performance on the QIYAS test is a good predictor of performance in the College of Education?

18/ If you could make changes to the way High School students are assessment, what changes would there be and why?

19/ Can you tell me anything else about the use of assessment in High School and in the College?

Appendix C   **Interview protocol after pilot**

study

The research title: The predictive validity of the admission standards in the college of education at King Faisal University in Saudi Arabia.

Date: **Place:**

Interviewer: Waleed Alshammari

Interviewee: **Position of interviewee:**

Briefly describe the study:

The purpose of the study is to investigate the predictive validity of the current admission standards applied at the College of Education at King Faisal University (KFU) in Saudi Arabia and explore which score among the current criteria used has the strongest contribution to students' academic success. Further, since this study attempts to include students from both gender groups, and very few studies have included both genders at the general level and to the best of my knowledge non has been done in the Saudi context. Before in the investigation, this study aims to explore any possible variation between the criteria items in terms of gender group. Also, since none of the previous studies have addressed the issue if students changing their major after being admitted to certain major initially at the university, this study attempt to explore the academic performance of students who changed their major after starting their university study.

Interview questions:

1/ What modules in the different year groups do you teach?

2/ Can you describe the different types of evaluation that we use in the College of Education?

3/ What type of evaluation do you use in your classes?

4/ What is the reasons of using this type of evaluation? Could you explain this for me?

5/ What do you know about the types of evaluation that use in the end of High School? Tell me as much as you can?

6/ What skills / knowledge do these assess?

7/ Can you give me some ways in which we are trying connect (align) the evaluation method of the High School with the evaluation with in the college?

8/ In your view, which type of admission standards that we use in the Education College has a higher relationship with the Grade Point Average (GPA) in the College of Education?

9/ Have you been able to predict the performance of student of knowledge of their QIYAS score?

10/ In your opinion, is that an accurate prediction? Can you give me some examples?

11/ Do you think that, a student who change their subject in the first year of their study may effect on their GPA? Can you talk more of that?

12/ As you know, QIYAS used multiple-choice but in Education College we use different types of evaluation. In your opinion, what should we doing in the College of Education to help student to improve their ability to do other types of evaluation?

13/ What skills are tested through the evaluation we use in Education?

14/ What knowledge is tested though the evaluation we use in Education?

15/ What links are there between evaluation at the end of High School and those in the College of Education?

16/ Do you think that performance on the QIYAS test is a good predictor of performance in the College of Education?

17/ If you could make changes to the way High School students are evaluation, what changes would there be and why?

18/ Do you want to add anything about the use of the evaluation in High School and in the college?

Appendix D   Interview question:

translate to Arabic

بروتوكول المقابلة الشخصية

عنوان الدراسة: القيمة التنبؤية لمعايير القبول المستخدمة في كلية التربية بجامعة الملك فيصل بالهفوف في المملكة العربية السعودية.

المكان: التاريخ:

الباحث: وليد بن غازي الشمري

المقابَل:

وظيفة المقابَل:

وصف مختصر للبحث: الغرض من الدراسة الحالية هو التحقق من الصدق التنبؤي لمعايير القبول الحالية المطبقة في كلية التربية في جامعة الملك فيصل في المملكة العربية السعودية ومعرفة أي معيار من بين المعايير الحالية المستخدمة (الاختبار التحصيلي، اختبار القدرات العامة وشهادة الثانوية العامة) لديه مساهمة أقوى في النجاح الأكاديمي لطلاب وطالبات كلية التربية. لا توجد أي دراسة سابقة في المملكة العربية السعودية ـحسب علم الباحثـ مماثلة شملت الإناث في عينتها، كما تهدف هذه الدراسة أيضا الى اكتشاف الأداء الأكاديمي للطلاب الذين غيروا تخصصهم بالانتقال الى كلية التربية من كلية اخرى في السنة الجامعية الاولى.

أسئلة المقابلة:

1/ منذ متى و انت تعمل في التعليم العالي؟

2/ ماهي المواد الدراسية التي تقوم بتدريسها في مختلف المراحل؟

3/ هل من الممكن ان تصف لي انواع الاختبارات وطرق التقويم التي تستخدمونها في كليتكم؟

4/ ماهي طرق التقويم التي تستخدمها انت في محاضراتك؟

5/ ماهي الاسباب التي دفعتك لاختيار هذه الطرق والاساليب؟ هل من الممكن ان تشرح اكثر؟

6/ ماذا تعرف عن اساليب التقويم المستخدمة في نهاية المرحلة الثانوية؟ اخبرني كل ما تعرف بإسهاب تفضلاً.

7/ ماهي المهارات و المعارف التي تقيسها هذه الطرق و الاساليب؟

8/ هل من الممكن ان تقترح لي طرقا تمكَّن من ربط اساليب التقويم المستخدمة في المرحلة الثانوية بتلك المستخدمة في الكلية؟

9/ من وجهة نظرك، أي معايير القبول التي تستخدم في كلية التربية بجامعة الملك فيصل لها علاقة أكبر بالمعدل التراكمي للطلاب في كلية التربية؟

10/ هل تستطيع توقع مستوى أداء و معرفة الطلاب من خلال درجات اختبارات مركز قياس؟

11/ في رأيك, هل هذا التوقع دقيق؟ هل من الممكن ان تعطي امثلة؟

12/ هل تعتقد ان تغيير التخصص في السنة الاولى قد يؤثر على مستوى الطالب الدراسي او معدله التراكمي؟ اشرح اكثر تفضلا.

13/ كما تعلم, اختبارات المركز الوطني للقياس والتقويم (قياس) تستخدم اسئلة الاختيار من متعدد, و في كلية التربية تُستخدم طرق اخرى. من وجهة نظرك ماذا يجب ان تقوم به كلية التربية حتى تساعد الطلاب في تطوير مهاراتهم في انماط التقويم الاخرى المستخدمة في الكلية؟

14/ ماهي المهارات التي يتم اختبارها او قياسها من خلال طرق التقويم المستخدمة في الكلية؟

15/ ماهي المعارف التي يتم اختبارها او قياسها من خلال طرق التقويم المستخدمة في الكلية؟

16/ ماهي الروابط بين طرق التقويم المستخدمة في المرحلة الثانوية والمستخدمة في كلية التربية؟

17/ هل تعتقد ان درجة الاداء في اختبارات (قياس) معيار جيد لتوقع اداء الطلاب في كلية التربية؟

18/ اذا كنت تستطيع تغيير الطريقة التي يتم فيها تقويم طلاب المرحلة الثانوية، ماهي التغييرات التي تقترحها؟ و لماذا؟

19/ هل تود اضافة أي شيء اخر عن طرق التقويم المستخدمة في المرحلة الثانوية و المستخدمة في كلية التربية؟

Appendix E : The assessment criteria used at King Faisal University

Arabic symbol	English symbol	Grade limits	points	Meaningful in Arabic	Meaningful In English
أ+	A+	100-95	5.00	ممتاز مرتفع	Exceptional
أ	A	94-90	4.75	ممتاز	Excellent
ب+	B+	89-85	4.50	جيدجداً مرتفع	Superior
ب	B	84-80	4.00	جيدجدا	Very good
ج+	C+	79-75	3.50	جيد مرتفع	Above average
ج	C	74-70	3.00	جيد	Good
د	D+	69-65	2.50	مقبول مرتفع	High-pass
د+	D	64-60	2.00	مقبول	Pass
هـ	F	59-0	1.00	راسب	Fail
م	IP	-	-	مستمر	In progress
ل	IC	-	-	غير مكتمل	Incomplete
ح	DN	-	-	محروم	Denial
ند	NP	60 or more	-	ناجح دون درجة	No grade-pass
هد	NF	Less than 60	-	راسب دون درجة	No grade-fail
ع	W	-	-	منسحب بعذر	Withdrawn
غم	US	-	-	غير مرضي	unsatisfactory

Appendix F: Correlation coefficient for the relationship between SGPA and UGPA

Correlations			
		High School GPA(SGPA)	University College GPA (UGPA)
High School degree	Pearson Correlation	1	.562**
	Sig. (2-tailed)		0.000
	N	693	693
**. Correlation is significant at the 0.01 level (2-tailed).			

Appendix G: Spearman's Correlation coefficient: General Aptitude Test (GAT) and UGPA

				General Aptitude Test (GAT)
Spearman's rho	Cumulative University Score (UGPA)	Correlation Coefficient		.324[**]
		Sig. (2-tailed)		.000
		N		693
		Bootstrap[b]	Std. Error	.033
			BCa 95% Confidence Interval	Lower	.260
				Upper	.382
a. **. Correlation is significant at the 0.01 level (2-tailed).				
b. Unless otherwise noted, bootstrap results are based on 1000 bootstrap samples				
c. BCa indicates the Bias Corrected Confidence Interval.				

Appendix H: Correlation coefficient: Achievement Test score (ACT) and UGPA

			Achievement Test score (ACT)
Cumulative University Score (UGPA)	Pearson Correlation		.268**
	Sig. (2-tailed)		.000
	N		384
		Std. Error	.058
		95% Bca. Confidence Interval — Lower	.157
		95% Bca. Confidence Interval — Upper	.385
a. **. Correlation is significant at the 0.01 level (2-tailed).			
b. Unless otherwise noted, bootstrap results are based on 1000 bootstrap samples			
c. Bca. indicates the Bias Corrected Confidence Interval			

Appendix I: Model summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.562^a	.315	.314	.773513	.315	318.310	1	691	.000
2	.583^b	.340	.338	.760012	.025	25.769	1	690	.000
3	.583^c	.340	.337	.760547	.000	.029	1	689	.865
a. Predictors: (Constant), End of High School assessment score (SGPA)									
b. Predictors: (Constant), End of High School assessment score (SGPA), General Aptitude Test (GAT)									
c. Predictors: (Constant), End of High School assessment score (SGPA), General Aptitude Test (GAT), Achievement Test score (ACT)									

Appendix J: Model fit statistics

ANOVA[a]						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	190.452	1	190.452	318.310	.000[b]
	Residual	413.441	691	.598		
	Total	603.893	692			
2	Regression	205.336	2	102.668	177.744	.000[c]
	Residual	398.556	690	.578		
	Total	603.893	692			
3	Regression	205.353	3	68.451	118.339	.000[d]
	Residual	398.540	689	.578		
	Total	603.893	692			
a. Dependent Variable: Cumulative University Score (UGPA)						
b. Predictors: (Constant), End of High School assessment score (SGPA)						
c. Predictors: (Constant), End of High School assessment score (SGPA), General Aptitude Test (GAT)						
d. Predictors: (Constant), End of High School assessment score (SGPA), General Aptitude Test (GAT), Achievement Test score (ACT)						

Appendix K: Regression coefficients

	Coefficients^a					
					95.0% Confidence Interval for B	
Model		Standardized Coefficients	T	Sig.	Lower Bound	Upper Bound
1	(Constant)	-3.207	-8.453	.000	-3.952	-2.462
	SGPA	.562	17.841	.000	.066	.082
2	(Constant)	-4.126	-9.956	.000	-4.940	-3.313
	SGPA	.512	15.795	.000	.059	.076
	GAT	.165	5.076	.000	.012	.028
3	(Constant)	-4.177	-8.173	.000	-5.181	-3.174
	SGPA	.512	15.777	.000	.059	.076
	GAT	.163	4.712	.000	.012	.028
	ACT	.006	.170	.865	-.009	.011
a. Dependent Variable: Cumulative University Score (UGPA)						

Appendix L: Mean scores by gender

Group Statistics					
Gender		N	Mean	Std. Deviation	Std. Error Mean
SGPA	Male	309	85.15	5.25	0.30
	Female	384	96.00	4.00	0.20
GAT	Male	309	72.33	6.73	0.38
	Female	384	77.81	7.49	0.38

Appendix M: Welch's robust test for equality of variance within groups

End of High School assessment score (SGPA)			
	Statistic[a]	df1	df2	Sig.
Welch	897.906	1	564.253	.000
a. Asymptotically F distributed.			

Appendix N: T-test results for gender differences in SGPA and GAT scores

Independent Samples Test

| | | Levene's Test for Equality of Variances | | t-test for Equality of Means | | | | | | |
| | | | | | | | | | 95% Confidence Interval of the Difference | |
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	S.E	Lower	Upper
End of High School assessment score (SGPA)	Equal variances assumed	37.27	0.00	-30.84	691.00	0.00	-10.84	0.35	-11.53	-10.15
	Equal variances not assumed			-29.97	564.25	0.00	-10.84	0.36	-11.55	-10.13
General Aptitude Test (GAT)	Equal variances assumed	1.62	0.20	-10.00	691.00	0.00	-5.47	0.55	-6.55	-4.40
	Equal variances not assumed			-10.12	682.66	0.00	-5.47	0.54	-6.54	-4.41

Appendix O: Mean differences by change of major

Change Major		N	Mean	Std. Deviation
End of High School assessment score (SGPA)	No change major	641	90.9	7.1
	Changed Major	52	94.1	6.3
General Aptitude Test (GAT)	No change major	641	78.8	6.9
	Changed Major	52	71.2	7.2
Achievement Test (ACT)	No change major	641	81.4	7.0
	Changed Major	52	69.9	6.9

Appendix P: Test of Homogeneity of Variances

|	Levene Statistic	df1	df2	Sig.
End of High School assessment score (SGPA)	7.693	1	691	.006
General Aptitude Test (GAT)	.140	1	691	.709
Achievement Test score (ACT)	.143	1	382	.706

Appendix Q: Independent Samples T-test for Change of Major

|		Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
End of High School assessment score (SGPA)	Equal variances assumed	7.69	0.01	-3.10	691.00	0.00	-3.14	1.02	-5.14	-1.15
	Equal variances not assumed			-3.44	62.12	0.00	-3.14	0.91	-4.97	-1.32
General Aptitude Test (GAT)	Equal variances assumed	0.14	0.71	4.10	691.00	0.00	4.48	1.09	2.34	6.62
	Equal variances not assumed			4.27	60.48	0.00	4.48	1.05	2.38	6.57
Achievement Test score (ACT)	Equal variances assumed	0.14	0.71	10.97	382.00	0.00	11.47	1.05	9.42	13.53
	Equal variances not assumed			11.12	68.64	0.00	11.47	1.03	9.41	13.53

Independent Samples Test

Appendix R: Scatter plot for the relationship between High School grade point average and university grade point average
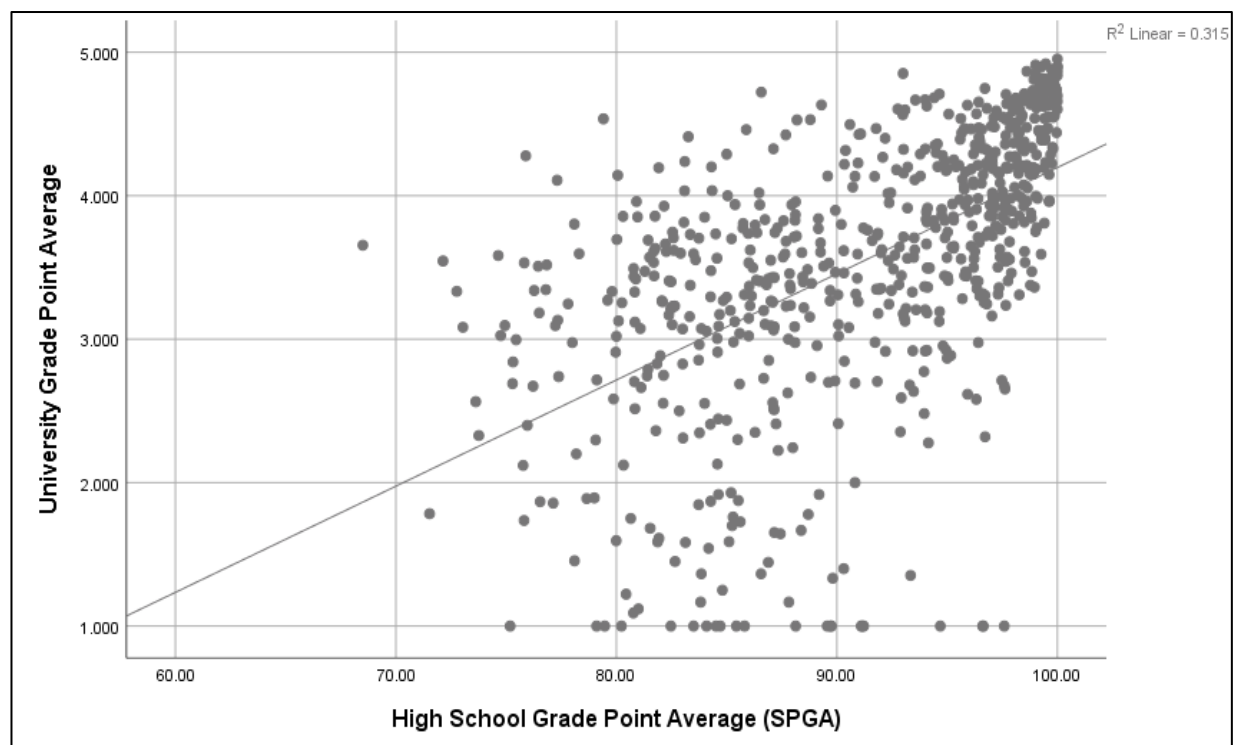
Appendix S: Scatter plot for the relationship between

**University Grade Point Average and General Aptitude Test**
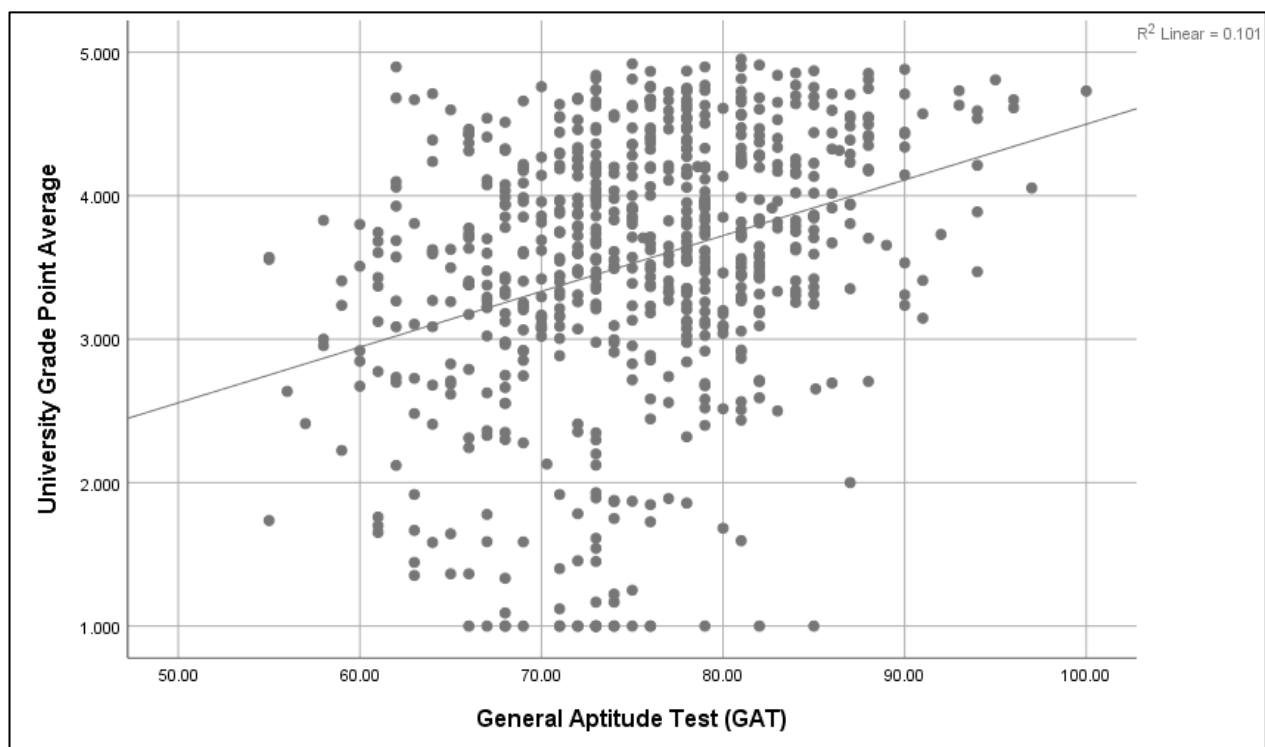
# Appendix T: Scatter plot for the relationship between University Grade Point Average and Achievement Test
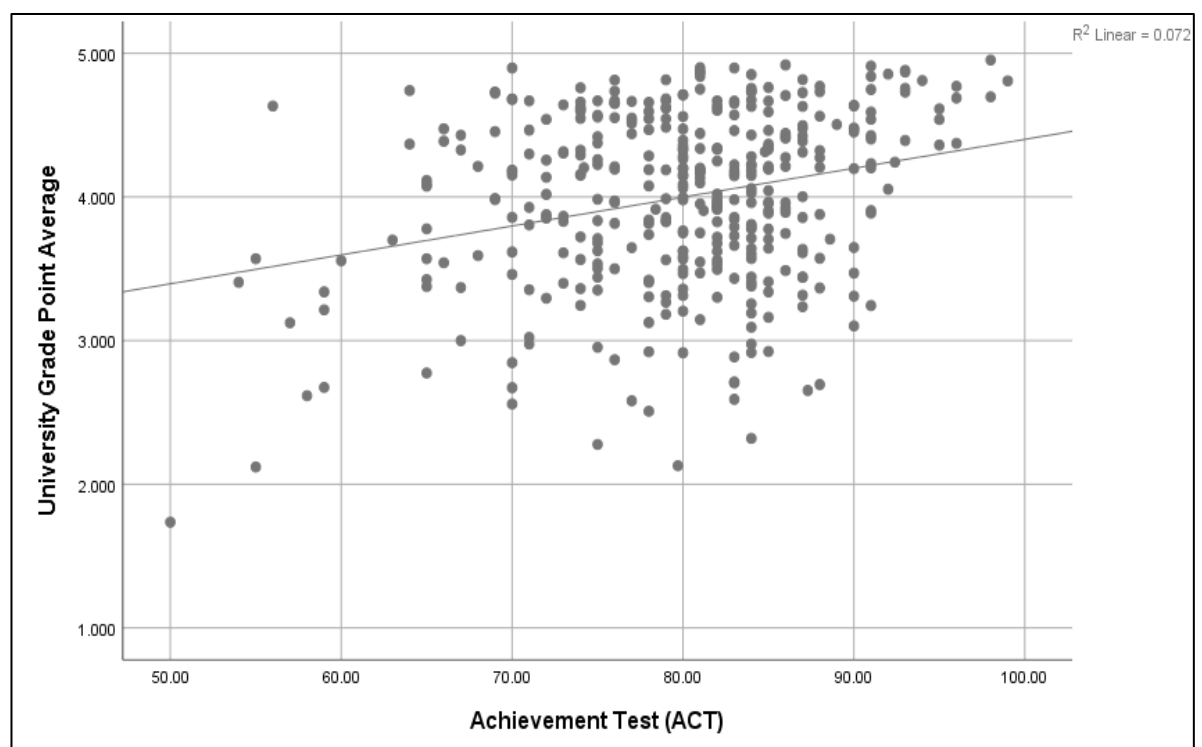
Appendix U: Histogram illustrating the assumption of normal distribution of residual

Histogram

Dependent Variable: GPA

Mean = -8.10E-1
Std. Dev. = 0.998
N = 693
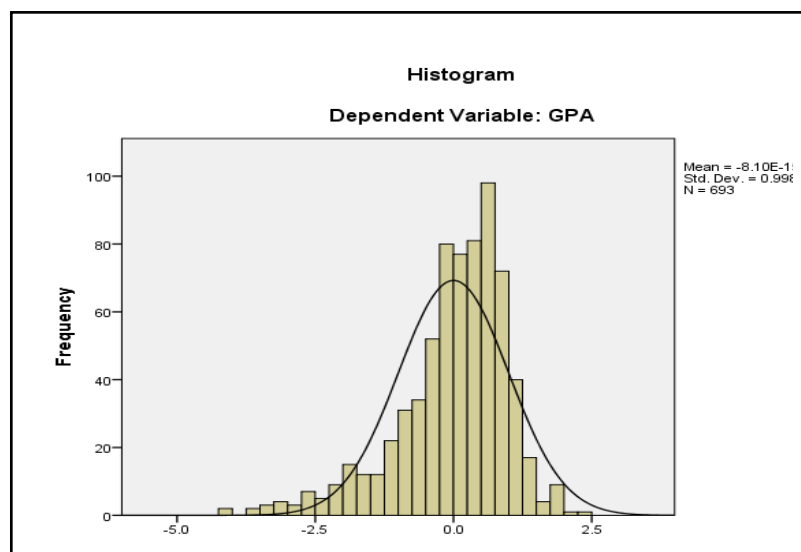
Appendix V: Normal P-P plot testing for normal distribution of residuals

Dependent Variable: University Grade Point Average
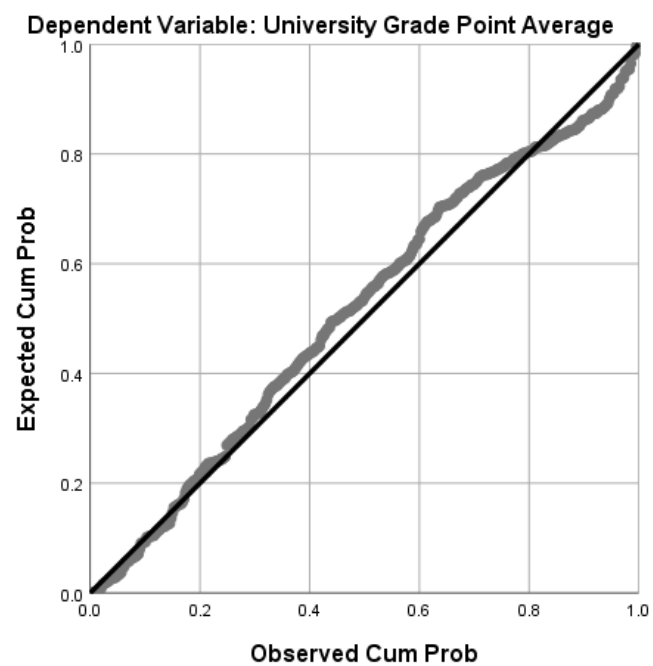
Appendix W: Scatter plot testing for constant variance

Dependent Variable: University Grade Point Average
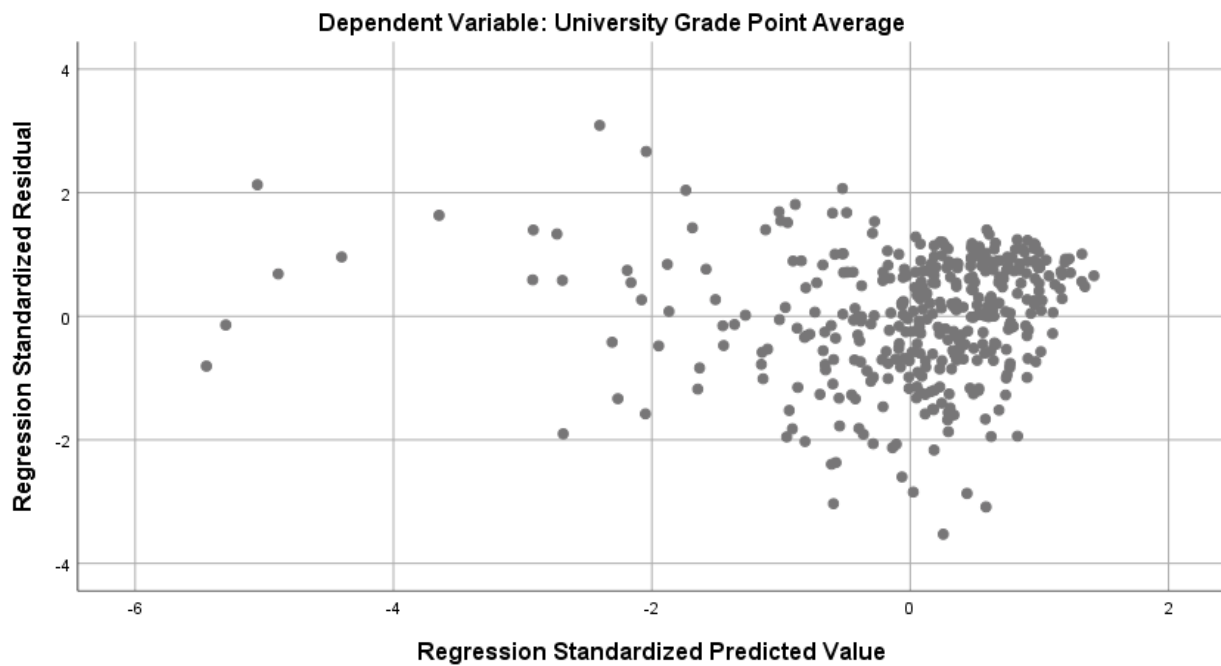
List of References

AERA, APA, & N. (2014). *Standards for educational and psychological testing*. Washington, DC: U.S.: American Educational Research Association.

Ahmed, A. (2008). *Ontological, Epistemological and Methodological Assumptions: Qualitative versus Quantitative*.

Al-Ahmadi, H. (2011). Challenges facing women leaders in Saudi Arabia. *Human Resource Development International*, *14*(2), 149–166. https://doi.org/10.1080/13678868.2011.558311

Al-Amri, A. (2014). The requirements of internationalization of higher education as an entry point to achieve the global leadership of Saudi universities, "Conceive a proposal", Ph.D. thesis, Department of Educational Administration and Planning, College of Education, Umm Al-Qura University. Saudi Arabia

Al-Asqah, H. & Dolat, A. (2017). The degree of using intermediate school teachers for alternative assessment methods in science teaching in the Qassim region in the Kingdom of Saudi Arabia, *Journal of Educational Sciences,* University of Jordan, 43 (1), 37-48

Al-Badri, S. M., & Al-Kindi, S. B. (2019). The reasons for the superiority of females over males from the point of view of those concerned in the field of education and the parents of students. *Rawafed Journal,*3 (1),71-115

Al-Dossary, S. (2001). *Predicting success in Higher Education:: predictive validity, attainment at school level and its relationship to degree class*. Durham University. Retrieved from http://etheses.dur.ac.uk/1533/1/1533.pdf

Al-Ghamdi, H & Gawad, N. (2015). *The development of the education system in Saudi Arabia*. Riyadh, Saudi Arabia: Rushd Bookstore.

Al-Ghanbousi, S. (2015), a proposed vision to activate the partnership between higher education institutions and sectors of society in the Arab Gulf states in light of the experiences of some international universities. *Working paper presented to the first international educational conference: developing the academic performance of colleges of education*: a forward-looking vision February 24-25, 2015, Al-Jouf University - Saudi Arabia

Al-Hajili, M. (2016). Degree of satisfaction and use of teachers in the Kingdom of Saudi Arabia for alternative assessment methods in evaluating the teaching process in general education and the obstacles to its use. *Journal of Education Science*, *24*(2), 206–261.

Al-Hattami, A. (2012). *Differential predictive validity of high school GPA and college entrance test scores for university students in Yemen*. University of Pittsburgh. Retrieved from http://d-scholarship.pitt.edu/11506/1/Abdu's__ETD_Dissertation.pdf

AL-Hazmi, S. (2003). EFL Teacher Preparation Programs in Saudi Arabia: Trends and Challenges, *TESOL Quarterly*, 37(2), 341-351.

Al-Kirani, M. (2010). T*he strategy of internationalizing university education in the Kingdom of Saudi Arabia for the post-accession phase of the World Trade Organization*, PhD thesis, Department of Educational Administration, College of Education, King Saud University, Riyadh. Saudi Arabia

Al-Malki, M. A. S. (2014). Testing the Predictive Validity of the IELTS Test on Omani English Candidates' Professional Competencies. *International Journal of Applied Linguistics & English Literature*, *3*(5), 166–172. https://doi.org/10.7575/aiac.ijalel.v.3n.5p.166

Al-Miziny, H. (2010). *The Abducting of Education in Saudi Arabia*.1st ed. Arab Diffusion Company, Beirut

Al-Omro, F & Degri, A. (2017). Attitudes and Perceptions of Saudi EFL Teachers in Implementing Kingdom of Saudi Arabia's Vision 2030. In *The role of Saudi's universities in activating Vision 2030* (pp. 810–843). Alqassim, Saudi Arabia: Alqassim University.

Al-Otaibi, K. (2006). *Saudi Higher Education's response to the challenges of globalizatio*n. PhD, Department of Educational Administration and Planning, College of Education: Umm Al-Qura University. Saudi Arabia.

Al-Sadhan, A. F. (2014) Appraisal of Quality Assurance Standards and Academic Accreditation on Marketing Education Outputs in Saudi Universities, T*he Fourth International Arab Council for Quality Assurance in Higher Education* (April 1-3, 2014) Zarqa University: Jordan

Abdul Qadir, O. (2017). The unified matrix of global rankings of Saudi universities on competitiveness in global indicators. *In The role of Saudi's universities in activating Vision 2030* (pp. 434--466). Alqassim, Saudi Arabia: Alqassim University

Alaam, S. (2005). *Educational and psychological measurement and evaluation fundamentals, applicationsand Contemporary orientations* (2ed ed., Vol. 2end). Cairo, Egypt: Dar Arab Thought.

Alacapinar, F. (2008). Effectiveness of project-based learning. *Egitim Arastirmalari - Eurasian Journal of Educational Research*, (33), 17–34.

Alameri, A. (2017). Building academic partnerships for educational graduate programs in Saudi universities in the light of the models of internationalization of Higher Education. In *Conference of the Role of Saudi Universities in Activating Vision 2030* (pp. 106–141). Al-Qassim, Saudi Arabia: Al-Qassim University.

Alaqeel, A. (2013). *Education policy and system in the Kingdom of Saudi Arabia*. Riyadh: Alroshed.

Alavi, T. (2012). The Predictive Validity of Final English Exams as a Measure of Success in Iranian National University Entrance English Exam. *Journal of Language Teaching and Research*, *3*(1), 224–228. https://doi.org/10.4304/jltr.3.1.224-228

Albanai, N., Balhadi, W., & Alkhawli, M. (2004). Predictive value of the admission criteria used in the University of Qatar and its relationship with the university cumulative average. *Educational Research Center,* Qatar University, 223.

Aldoghan, A. (1985). *The predictive validity of selection measures used by the University of Petroleum and Minerals in Saudi Arabia.* Michigan State University.

Alfassi, H. (2006). *A Future Vision For The Private Higher Education Sector in The Gulf Cooperation Council Countries (GCC).* Umm Al Qura University.

Algarfi, A. (2010). *Teachers and pupils perceptions of and responses to cooperative learning methods within the islamic culture courses in one secondary school in saudi arabia*. UNIVERSITY OF SOUTHAMPTON.

Alghamdi, H & Abduljawad, N. (2005). *The development of the education system in the Kingdom of Saudi Arabia*. Riyadh: Alroshed.

Alghamdi, M. (2007). T*he predictive value of the general aptitude test and the general secondary school standard as admission criteria for the students to join Umm-Alqura University*. PhD, College of Education: University of Umm-Alqura- Saudi Arabia

Alhajri, F. S. M. (2013). *An Evaluation of the Effectiveness and Predictive Validity of English Language Assessment in Two Colleges of Applied Sciences in Oman*. The University of Edinburgh.

Alhammed, M. (2005). *The Development of Education in Saudi Arabia*. Riyadh: Alroshed.

Alhammed, M., Zeadah, M., Alotaiby, B & Motawaly, N. (2007). *The learning In Saudi Arabia "sight of the present and preparing for future*. Riyadh: Alroshed.

Alharbi, A. (2011). *The Development and Implementation of a CPD Programme for Newly Qualified Teachers In Saudi Arabia*. University of Southampton. Retrieved from http://eprints.soton.ac.uk/209275/1/PHD_theses_A.Alharbi.pdf

Alhogail, S. (2011). *The Law and Strategies of Learning in Saudi Arabia*. Riyadh: Alroshed.

Alhojailan, M. (2012). Thematic Analysis : A Critical Review of Its Process and Evaluation. *West East Journal of Social Sciences*, *1*(1), 39–47.

Ali, A. (2017). Investing in Higher Education and improving its outputs to achieve the objectives of Vision 2030. In *The role of Saudi's universities in activating Vision 2030* (pp. 636–695). Alqassim, Saudi Arabia: Alqassim University, Saudi Arabia.

Aljassar, S. (2012). Assess the general conditions for admission to the colleges of education in the Arab Gulf states. *Journal of College of Education - Tanta University-Egypt*, *51*(2), 215–238.

Aljuraid, A. B. (2013). *Holy Quran Memorization Schools in the Kingdom of Saudi Arabia, a bright face in the education system*.Riyadh

Alkhalidi, H & Alturki, O. (2018). Impact of effective feedback in learning management systems on enhancing student learning outcomes. *International Interdisciplinary Journal of Education*, *7*(7), 115–129.

Alkhattab, A. (1989). Open admission for Kuwaiti students and their achievement in college. In *Annual Conference eleventh - Arab Organization responsible for Admission to universities in Arab countries*. Amman: University of Jordan in Amman.

Alkushi, A & Althewini, A. (2020). The Predictive Validity of Admission Criteria for College Assignment in Saudi Universities: King Saud bin Abdulaziz University for Health Sciences Experience. I*nternational Education Studies*. 13. (4), 141.

Allwood, C. M. (2012). The distinction between qualitative and quantitative research methods is problematic. *Quality & Quantity*, 46(5), 1417-1429.

Almarshad, A. (1996). *Criteria for accepting high school students in universities*. Riyadh.

Almiqdadi, F. (2010). The relationship between high school GPA, math and calculus 1 & 2 scores and the GPA of the UAE University female students. *University of Sharjah Journal Humanities and Social Sciences*, *7*(2), 65–89.

Almokhlafi, M. (2001). The effectiveness of scondary school rate in predicting academic achievement for students in the College of Education in Amran - Sana'a University. *Journal of Educational and Psychological Sciences, Damascus University*, *17*(4), 209–235.

Alnahdi, G. H. (2013). Educational Change In Saudi Arabia. *Journal of International Education Research (JIER)*, *10*(1), 1. https://doi.org/10.19030/jier.v10i1.8342

Alnahdi, G. H. (2015). Aptitude Tests and Successful College Students: The Predictive Validity of the General Aptitude Test (GAT) in Saudi Arabia. *International Education Studies*, *8*(4), 1–6.

https://doi.org/10.5539/ies.v8n4p1

Alnajjar, A. (2001). The Predictive Validity Of The Admission Standards At King Faisal University. *Educational Journal- Arabic*, *15*(59), 217.

Alomair, M. (2018). *Collegiate women in Saudi Arabia : Leading collectively for the development of felf,others, and society*. Chapman University,California.

Alqatei, A., & Alharbi, K. (2012). *The predictive validity of the univeisity admission standards in the GPA of first year students at some Saudi University*. Riydh.

Alsaif, A. (2005). *The Predictive validity of the admission standards at King Fahd University of Petroleum & Minerals*. King Saudi University.

Alsalem, A. (2017). Developing proposed standards for attracting, preparing and training teachers in the Kingdom of Saudi Arabia in the light of Vision 2030. In *The role of universities in activating Vision 2030* (pp. 320–355). Al Qussaim Saudi Arabia: Al Qussaim university.

Alsaqri, A., & Almuhaimed, Y. (2017). Training evaluation as an input to enhance the competitiveness of Saudi universities in the global indicators. *In The role of Saudi's universities in activating Vision* 2030 (pp. 467–498). Alqassim, Saudi Arabia: Alqassim University.

Alshamlan, K and Al-Fawzan, A. (2017). The requirements of universities to implement strategic management in order to achieve the vision of 2030 from the perspective of the members of the Saudi Management Association. In *Conference of the Role of Saudi Universities in Activating Vision 2030* (pp. 229–268). Al-Qassim, Saudi Arabia: Al-Qassim University.

Alshehri, A. (2011). The Predictive Validity Of The Admission Standards At AL-Taif University. In *In 31 Annual Conference3 meeting- Arab League for admission Administrators at Arabic Universites*. Abu Dhabi UAE: Alhisn University.

Alshumrani, S. (2007). *Predictive validity of the General Aptitude Test and High School Percentage for Saudi undergraduate students*. University of Kansas. Retrieved from http://gradworks.umi.com/32/66/3266513.html

Alsunbol, A., Alkateep, M., Motoaly, M & Abduljawad, N. (2009). S*ystem of Education in Kingdom of Saudi Arabia*. 2nd ed. Riyadh: Daar Alkhareji.

Alyahyan, E., & Düştegör, D. (2020). Predicting academic success in higher education: literature review and best practices. *International Journal of Educational Technology in Higher Education,* 17(1), 3.

Alzahrani, B., H. (2000). *Predictive Validity of admission Criteria used at teachers Colleges in Saudi Arabia*. Master's Thesis. College of Education. The University of Umm Al-Qura.

Alzaki, A. (2017) University Services Marketing: An urgent necessity to enhance the financial resources of Saudi universities in light of Vision 2030. I*n The role of Saudi's universities in activating Vision 2030* (pp. 606–632). Alqassim, Saudi Arabia: Alqassim University.

Amer, M. (2015). *Project-Based Learning*. Sohar, Oman.

Attia, N. (2002). *Entrance exams at King Fahd University of Petroleum and Minerals .pdf*. Riyadh. Retrieved from http://gesten.ksu.edu.sa/en

Au, W. (2007). High-Stakes Testing and Curricular Control: A Qualitative Metasynthesis. *Educational Researcher*, *36*(5), 258–267. https://doi.org/10.3102/0013189X07306523

Azizi, I. (2018). Effectiveness of Formative Assessment in Improving Academic Achievement among Students of Thinking and Scientific Research Skills Course, Faculty of Business Administration, Shaqra University. *Journal of the College of Basic Education for Educational and Human Sciences, Babylon University*, *41*(10), 678–711.

Bachman, L. (2000). Modern language testing at the turn of the century: Assuring that what we count counts. *Language Testing – Lang test*. 17. 1-42.

Bagozzi, R. P., Yi, Y., & Phillips, L. W. (1991). Assessing construct validity in organizational research. *Administrative science quarterly*, 421-458.

Baker, T. L. (1994). *Doing Social Research* (2nd ed.). New York: McGraw-Hill Inc. https://doi.org/10.1017/CBO9781107415324.004

Balta, V. 2020, UW removes standardized testing requirement for incoming students beyond fall 2021, washington Education viewed 16 September 2020,<https://www.washington.edu/news/2020/06/11/uw-removes-standardized-testing-requirement-for-incoming-students-beyond-fall-2021/>

Bany Odeh, K. (2015). *The effect of using the alternative assessment on the collection of ninth grade students and their attitudes towards science in Nablus schools.* An-Najah National University, Palestinian.

Barnes, N., Fives, H., & Dacey, C. M. (2015). Teachers' beliefs about assessment. *International Handbook of Research on Teacher Beliefs*. https://doi.org/10.1007/s13398-014-0173-7.2

Barrett, H. (2006). Using electronic portfolios for formative/classroom-based assessment. *Classroom Connect Connected Newsletter*, *13*(2), 4–6. Retrieved from https://electronicportfolios.org/portfolios/ConnectedNewsletter.pdf

Bassey, M. (1999). *Case Study Research in Educational Settings*. Buckingham, Open University Press.

Bati, T., Gelderblom, H., & van Biljon, J. (2014). *Blended learning of programming in large classes a reflection of students' experience from an Ethiopian University*. University of Cape Town.

Bennett, R. E. (2011). Formative assessment: a critical review. *Assessment in Education: Principles, Policy & Practice*, *18*(1), 5–25. https://doi.org/10.1080/0969594X.2010.513678

Bergmann, E. (2014). *An Examination of the Relationship between the Frequency of Standardized Testing and Academic Achievement*. University of Oregon.

Berrett, D. (2013, January 21). My GRE score says I'm smart. Hire me. *The Chronicle of Higher Education*.

Bhati, S. S. (2012). The effectiveness of oral presentation assessment in finance subject: An empirical examination. *Journal of University Teaching & Learning Practice*, *9*(2). Retrieved from http://ro.uow.edu.au/cgi/viewcontent.cgi?article=1187&context=jutlp

Billows, J. E. (2007). *Predictive validity of an innovative selective admissions process for baccalaureate nursing students*. PhD Thesis, University of British Columbia.

Bird, C. M. (2005). How I stopped dreading and learned to love transcription. *Qualitative Inquiry,* 11(2), 226-248.

Birgin, O., & Baki, A. (2007). The Use of Portfolio to Assess Student's Performance. *Journal of Turkish Science Education*, *4*(2), 75–90. Retrieved from http://www.tused.org/internet/tufed/arsiv/v4/i2/metin/tusedv4i2s6.pdf

Black, P. J. (1998). *Testing, friend or foe? the theory and practice of assessment and testing*. Psychology Press.

Black, P., & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 7–74. https://doi.org/10.1080/0969595980050102

Boeije, H. (2002). A Purposeful Approach to the Constant Comparative Method in the Analysis of Qualitative Interviews. *Quality and Quantity*, *36*(4), 391–409. https://doi.org/10.1023/A:1020909529486

Bollen, K. A. (1989). *Structural Equations with Latent Variables*. *Structural Equations with Latent Variables*. New York ; Chichester, US: John Wiley & Sons.

Bond, L. (1996). Norm and Criterion Referenced Testing. *Practical Assessment, Research & Evaluation*, *5*(2). https://doi.org/10.1037/13619-020

Borg, W.R. & Gall, M. D. (1996) *Educational Research: An Introduction* (sixth edition). New York: Longman.

Bouma, G. D. & Atkinson, G. B. (1995). *A Handbook of Social Science Research: A Comprehensive and Practical Guide for Students*. Oxford: Oxford University Press.

Bourner, T., & Hamed, M. (1987). degree performance: report to CNAA Development Services of a research project on the relationship between entry qualifications and degree performance on CNAA. *Development Services Publication*. Retrieved from http://www.voced.edu.au/word/43696

Boyatzis, R. (1998). *Transforming qualitative information: Thematic analysis and code development*. Thousand Oaks, CA: Sage.

Boyd, H. W., Westfall, R. L. & Stasch, S. F. (1985) Marketing research: text and cases, R.D. Irwin, 1985, Homewood, Ill.

Brannen, J. (1995). *Mixing methods: qualitative and quantitative research*.Ashgate: Aldershot: Avebury.

Brannen, J. (2005) "Mixed Methods research: A discussion paper", *ESRC National Centre for Research Methods NCRM Methods Review papers*, NCRM/005 [online] http://eprints.ncrm.ac.uk/89/1/MethodsReviewPaperNCRM-005.pdf

Braun, V. and Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Braun, V., & Clarke, V. (2013). *Successful qualitative research: A practical guide for beginners*. sage.

Brewi, E. (2015). *The Pros and Cons of Standardized Testing* (Vol. 1). Long Island University/Post. https://doi.org/10.1017/CBO9781107415324.004

Broadfoot, P. M. (1984). T*he rationality of judgement: comparative perspectives on the social role of educational assessment*. PhD thesis. The Open University.

Brookhart, S. M. (2008). Portfolio Assessment. In T. L. Good (Ed.), *21st Century Education : A Reference Handbook* (pp. 443–450). Thousand Oaks, California: Sage Publications Ltd.

Brown, J. D. (1990). *Testing in Language Program*s. New Jersey: Prentice Hall Regent.

Brown, J & Hudson, T. (1998). The Alternatives in Language Assessment. *TESOL Quarterly*, *32*(4),

653–675. https://doi.org/10.1017/CBO9781107415324.004

Brown, R. S., & Coughlin, E. (2007). The predictive validity of selected benchmark assessments used in the Mid-Atlantic Region. *Regional Educational Laboratory* (Issues & Answers Report, REL 2007–No. 017). Washington, DC: U.S.

Brown, GAVIN T. L. (2006). TEACHERS' CONCEPTIONS OF ASSESSMENT: VALIDATION OF AN ABRIDGED VERSION. *Psychological Reports*, *99*(5), 166. https://doi.org/10.2466/pr0.99.1.166-170

Brown, Gavin T.L., & Gao, L. (2015). Chinese teachers' conceptions of assessment for and of learning: Six competing and complementary purposes. *Cogent Education*, *2*(1), 1–19. https://doi.org/10.1080/2331186X.2014.993836

Brown, M., Burke, K., Curtis, P., Farrell, P., Focht, D., Georgianna, L., Moore, C. (2002). The Use of Admissions Tests by the University of California, A Discussion Paper Prepared by *the Board of Admissions and Relations with Schools*, 1–18.

Brown, S. P. (2011). *Assessment in the further education and training school sector: a quality assurance perspective*. University of the Free State,Bloemfontein.

Bryman, A. (2003). *Quantity and quality in social research*. Routledge.

Bryman, A. (2008). *Social research methods*. Oxford university press.

Buela-Casal, G., Gutiérrez-Martínez, O., Bermúdez-Sánchez, M. P., & Vadillo-Muñoz, O. (2007). Comparative study of international academic rankings of universities. *Scientometrics*, 71(3), 349-365.

Burns, A. (1999). *Collaborative action research for English language teachers*. Cambridge: CUP.

Camara, W., & Echternacht, G. (2000). *The SAT I and High School Grades: Utility in Predicting Success in College. Research Notes.* New York. Retrieved from http://files.eric.ed.gov/fulltext/ED446592.pdf

Camara, W. J. (2008). Score trend, SAT validity, and subgroup differences. *Paper presented at the Harvard Summer institute*, Boston.

Cameron, R. (2011). Mixed methods research: The five Ps framework. *Electronic Journal of Business Research Methods*, 9(2).

Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment*. Sage publications.

Carty, H., & Armstrong, W. (2008). The Predictive Validity f o the New SAT Achievement test and High School Grades on Undergraduate Success. In *Presented to the Annual Forum of the Association for Institutional Research* (pp. 1–22). San Diego, U.S.A: University of California, San Diego.

Chappius, S., Stiggins, R., Arter, J., & Chappius, J. (2009). *Assessment for learning : an action guide for school leaders* (2nd ed.). Portland: Pearson.

Chase, C. Z. (2009). *Measurement for educational evaluation*. Massachusetts: Addison-Wesley Publishing Company.

Chen, H. mei, Yu, C., & Chang, C. sian. (2007). E-Homebook System: A web-based interactive education interface. *Computers and Education*, *49*(2), 160–175. https://doi.org/10.1016/j.compedu.2005.05.003

Chioncel, N. E., Veen, R.G.W., Wildemeersch, D. & Jarvis, P. (2003) The Validity and Reliability of Focus Group as a Research Method in Adult Education. I*nternational Journal of Lifelong Education*, 22, (5), 495-517.

Chirimbu, S. (2013). Using Alternative Assessment Methods in Foreign Language Teaching . Case Study : Alternative Assessment of Business English for University Students. *Scientific Bulletin of the Politehnica University of Timişoara. Transactions on Modern Languages*, *12*(1), 91–98.

Cho, Y., & Bridgeman, B. (2012). Relationship of TOEFL iBT(R) scores to academic performance: Some evidence from American universities. *Language Testing*, *29*(3), 421–442. https://doi.org/10.1177/0265532211430368

Clarke, M. (2011). Framework for Building an Effective Student Assessment System. Read/Sabe*r Working Paper*.

Cohen, L., Manion, L. & Morrison, (2007). *Research Methods in Education* (5th ed.). London: Routledge Falmer.

Cooper, J. (n.d.). *Qualitative Research Design: An Interactive Approach* (Third). London: SAGE Publications. Retrieved from http://onlinelibrary.wiley.com/doi/10.1525/aeq.1998.29.4.499/full

Cooper, J., Brandon, P., & Lindberg, M. (1997). Using peer debriefing in the final stage of evaluation with implications for qualitative research: Three impressionist tales. *Paper presented at the American Educational Research Association*, Chicago, IL. (ERIC Document Reproduction Service No. ED410287)

Cotton, F., & Conrow, F. (1998). *An investigation of the predictive validity of IELTS amongst a group of international students studying at the University of Tasmania. IELTS research reports* (Vol. 1). Sydney:ELICOS Association Limited.

Creswell, J. W. (2007). *Qualitative inquiry and research design: Choosing among five traditions*. Thousand Oaks, CA: Sage Publications

Dann, R. (2014). Assessment as learning: Blurring the boundaries of assessment and learning for theory, policy and practice. *Assessment in Education: Principles, Policy & Practice*, *21*(2), 149–166. https://doi.org/10.1080/0969594X.2014.898128

Darandari, E. . (2010). Assessment of learning outcomes: Towards a new framework that utilizes contemporary trends for assessment and learning quality. *Paper Published by Research Centre in College of Education, King Saud University, Riyadh.*, 1–19.

Darling-Hammond, L. (2008). Powerful learning: What we know about teaching for understanding. *Powerful Learning: What We Know about Teaching for Understanding.* San Francisco, CA: Jossey-Bass. Retrieved from http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc6&NEWS=N&AN=2008-12930-000

Darwaza, A. (1987). *Predict academic achievement University of previous academic factors in student achievement*. University of Jordan, Amman.

Dayioglu, M., & Turut-Asik, S. (2007). Gender differences in academic performance in a large public university in Turkey. *Higher Education*, 53(2), 255-277.

Denzin, N. K., & Lincoln, Y. S. (2008). *Strategies of qualitative inquiry* (Vol. 2). Sage.

DeVon, H. a., Block, M. E., Moyle-Wright, P., Ernst, D. M., Hayden, S. J., Lazzara, D. J., … Kostas-

Polston, E. (2007). A psychometric toolbox for testing validity and reliability [Electronic Version]. *Journal of Nursing Scholarship*, *39*(2), 155–164. https://doi.org/10.1111/j.1547-5069.2007.00161.

De Wit, H., Hunter, F., Howard, L., & Egron-Polak, E. (2015). *Internationalisation ofHigher Education*. Milan: Brussels: Policy Department, Directorate General for Internal Policies, European Parliamen. https://doi.org/02-15-573-EN-N.

Dikli, S. (2003). Assessment at a distance : Traditional vs . Alternative Assessments. *The Turkish Online Journal of Education Technology*, *2*(3), 13–19. https://doi.org/ISSN 1303-6521

Dill, D. D. (2003). Allowing the market to rule: The case of the united states. *Higher Education Quarterly*, *57*(2), 136–157. https://doi.org/10.1111/1468-2273.00239

Dooey, P., & Oliver, R. (2002). An investigation into the predictive validity of the IELTS Test as an indicator of future academic success. *Prospect-Adelaide-*, *17*(1), 36–54.

Doumato, E. (2002). Women and Work in Saudi Arabia: How flexible are Islamic margins? *Middle East Journal*, 53 (4), 568-583

Drost, E. a. (2004). Validity and Reliability in Social Science Research. *Education Research and Perspectives*, *38*(1), 105–123.

Duckworth, A. L., & Seligman, M. E. (2006). Self-discipline gives girls the edge: Gender in self-discipline, grades, and achievement test scores. *Journal of educational psychology*, 98(1), 198-208

Earl, L. M. (2012). *Assessment as Learning: Using Classroom Assessment to Maximise Student Learning* (2nd editio). Thousand Oaks,Califonia: Corwin Press.

Eckes, T. (2008). Rater types in writing performance assessments: A classification approach to rater variability. *Language Testing*, 25(2), 155-185.

Edmunds, A. (2010). *An Examination of the Likelihood of Persistence of Students with Discrepant High School Grades and Standardized Test Scores*. The University of Alabama.

Educational Testing Services. (2015). *GRE Guide to the Use of Scores 2015-2016*.

Edwards J C; Johnson, E K; Molidor, J. B. (1990). The interview in the admission process. *Association of American Medical Colleges*, *65*(3), 167–177. https://doi.org/2407259

Elharrar, Y. (2006). *TEACHER ASSESSMENT PRACTICES AND PERCEPTIONS: THE USE OF ALTERNATIVE ASSESSMENTS WITHIN THE QUEBEC EDUCATIONAL REFORM*. Universite du Quebec a Montreal, Canada.

Elmulthum, N., & Elsayed, I. (2017). Prospects of Saudi Women ' s Contribution to Job Market under Saudi Vision 2030 : An Empirical Analysis. *International Journal of Applied Sociology*, *7*(1), 20–27. https://doi.org/10.5923/j.ijas.20170701.03

Elyas, T. (2008). The attitude of American English within the Saudi education system. *Novitas-Royal,* 2(1), 28–48.

Espinosa, M. P., & Gardeazabal, J. (2010). Optimal correction for guessing in multiple-choice tests. *Journal of Mathematical Psychology*, *54*(5), 415–425. https://doi.org/10.1016/j.jmp.2010.06.001

ETS. (2012). *Practice Book for the Paper-based GRE revised General Test* (Second Edi). United States: Educational Testing Service (ETS).

Field, A. (2013). *Discovering statistics using IBM SPSS statistic*s. sage.

Fives, H., & Buehl, M. M. (2012). Spring cleaning for the "messy" construct of teachers' beliefs: What are they? Which have been examined? What can they tell us? *APA Educational Psychology Handbook, Vol 2: Individual Differences and Cultural and Contextual Factors.*, *2*, 471–499. https://doi.org/10.1037/13274-019

Fleming, J., & Garcia, N. (1998). Are standardized test fair to African Americans? *Journal of Higher Education*, *6*(9), 471–495.

Foss, C. & Ellefsen, B. 2002. The value of combining qualitative and quantitative approaches in nursing research by means of method triangulation. *Journal of Advanced Nursing*, 40(2): 242-248.

Frank, M., & Barzilai, A. (2004). Integrating alternative assessment in a project-based learning course for pre-service science and technology teachers. *Assessment & Evaluation in Higher Education*, *29*(1), 41–61. https://doi.org/10.1080/0260293042000160401

Frankfort-Nachmias, C., & Nachmias, D. (2007). Study guide for research methods in the social sciences. Macmillan.

Frazier, D. M., & Paulson, F. L. (1992). How Portfolios Motivate Reluctant Writers. *Educational Leadership*, *49*(8), 62–65.

Furnham, A., & Monsen, J. (2009). Personality traits and intelligence predict academic school grades. *Learning and Individual Differences*, *19*(1), 28–33. https://doi.org/10.1016/j.lindif.2008.02.001

Gall, M. D., Borg, W. R., & Gall, J. P. (1996). *Educational research: An introduction*. Longman Publishing.

Gardner, J. (2012). *Assessment and Learning* (2nd ed.). London, SAGE Publications. Retrieved from http://www.uk.sagepub.com/upm-data/43829_9780857023834.pdf

Gawthrop, J. (2014). Measuring Student Achievement: A Study of Standardized Testing & Its Effect on Student Learning Measuring Student Achievement : *A Study of Standardized Testing and its Effect on Student Learning*, 44.

Geiser, S & Atkinson, R. (2009). Reflections on a century of College Admissions Tests. *Advances in Parasitology*, *67*(April), 1–402. https://doi.org/10.1016/S0065-308X(08)00405-3

Geiser, S, & Studley, R. (2001). UC and the SAT: Predictive validity and differential impact of the SAT I and SAT II at the University of California. *Educational Assessment*. Retrieved from http://www.tandfonline.com/doi/abs/10.1207/S15326977EA0801_01

Geiser, Saul, & Studley, R. (2010). UC and the SAT: Predictive validity and differential impact of the SAT I and SAT II at the University of California. *Educational Assessment*, *8*(1), 1–26. https://doi.org/10.1207/S15326977EA0801

Gill, J. T. (2007). *Fair admissions and fair access implications of assessing aptitude, personality and attitude in student selection for Higher Education*. University of Manchester.

Given, L. M. (Ed.). (2008). *The Sage encyclopedia of qualitative research method*s. Sage publications.

Glesne, C. & Peshkin, A. (1992). *Becoming qualitative researchers: An introduction*. Pearson. Longman, White Plains, N.Y.

Glickman, C. D., Gordon, S. P., & Ross-Gordon, J. M. (2009). *Supervision and instructional Leadership: A developmental approach*. Boston MA: Allyn and bacon.

Golafshani, N. (2003). Understanding reliability and validity in qualitative research. *Qualitative Report*, 8, 597-606. Retrieved July 2, 2018 from http://www.nova.edu/ssss/QR/QR8-4/golafshani.pdf

Gon, S., & Rawekar, A. (2017). Effectivity of E-Learning through Whatsapp as a Teaching Learning Tool. *MVP Journal of Medical Sciences*, *4*(1), 19–25. https://doi.org/10.18311/mvpjms/2017/v4i1/8454

Good, Th. L. (2008). *21st Century Education : A Reference Handbook*. Thousand Oaks, California: Sage Publications Ltd.

Grady, D. G., Cummings, S. R., & Hulley, S. B. (2013). Research using existing data. *In Designing clinical research* (4th ed., pp. 192–204). Philadelphia: Lippincott Williams & Wilkins*.*

Graham, C., & Neu, D. (2004). Standardized testing and the construction of governable persons. *Journal of Curriculum Studies*, *36*(3), 295–319. https://doi.org/10.1080/0022027032000167080

Gronlund, N. W., Linn, R. L., Gronlund, N. E., & Linn, R. L. (1990). Validity, reliability, and other desired characteristics. *Measurement and evaluation in teaching*. New York, NY: Macmillan, 47-106.

Gumbo, M. T. (2014). An Action Research Pilot Study on the Integration of Indigenous Technology in Technology Education. *Mediterranean Journal of Social Sciences*, *5*(10), 386–392. https://doi.org/10.5901/mjss.2014.v5n10p386

Hakim, A. (2012). *Education System and Polic*y. Cairo: Itrac for printing and publishing.

Harris, L., & Brown, G. (2010). Mixing interview and questionnaire methods: Practical problems in aligning data. *Practical Assessment Research & …*, *15*(1), 1–19. Retrieved from http://libir1.ied.edu.hk/dspace/handle/2260.2/10032

Hassan, A. M. A. H. (2006). *Psychometric properties of the measuring instruments in the psychological and educational research using SPSS*. Riyadh.

HEC. (2015). *Higher Education Council and universities System and regulations* (4th ed.). Riyadh, Saudi Arabia: King Fahd National Library.

Heilig, J. V., & Darling-Hammond, L. (2008). Accountability Texas-Style: The Progress and Learning of Urban Minority Students in a High-Stakes Testing Context. *Educational Evaluation and Policy Analysis*, *30*(2), 75–110. https://doi.org/10.3102/0162373708317689

Hemsley-Brown J.V. and Oplatka, I. (2010). Universities in a competitive global marketplace: a systematic review of the literature on Higher Education marketing. *International Journal of Public Sector Management*, *19*(4), 316–338. https://doi.org/doi:10.1108/09513550610669176

Herman, J. (1997). Assessing new assessments: How do they measure up? *Theory Into Practice*, *36*(4), 196–204. https://doi.org/10.1080/00405849709543769

Hidri, S. (2016). Conceptions of assessment : Investigating what assessment means to secondary and university teachers. *Arab Journal of Applied Linguistics*, *1*(1), 19–43.

Higher Committee for Education Policy. (1982). E*ducation Policy Document in the Kingdom of Saudi Arabia*, Riyadh

Hill, K., Storch, N., & Lynch, B. (1999). Comparison of IELTS and TOEFL as predictors of academic success. *International English Language Testing System (IELTS) Research Reports*, *2*, 63–73.

Hopper, C. H. (2012). *Practicing College Learning Strategies* (6 edition). Cengage Learning, Inc.

Huerta-Macias, A. (1995). Alternative assessment: Responses to commonly asked questions. *TESOL Journal*, 5(1): 8–11.

Iipinge, S. M. (2013). *Challenges of Large Class Teaching at the University: Implications for Continuous Staff Development Activities*.

Imam Muhammad Bin Saud Islamic University. (1991). C*urricula of Secondary Stage in Scientific Institutes*. Riyadh

Janisch, C., Liu, X., & Akrofi, A. (2007). Implementing alternative assessment: opportunities and obstacles. *The Educational Forum*, 71 (3), 221-230.

Kaiser, J. (1982). The predictive validity of GRE aptitude test. *Paper presented at the annual meeting of the Rocky Mountain Research Association*, Albuquerque, NM. (ERIC Document Reproduction Service No. ED 226 021)

Kaneko, M. (2007). *Marketization of Higher Education Trends Issues and Prospects*. Tokyo, Japan.

Kass, D., Grandzol, C., & Bommer, W. (2012). The GMAT as a predictor of MBA performance: Less success than meets the eye. *Journal of Education for Business*, *87*(5), 290–295. https://doi.org/10.1080/08832323.2011.623196

Kavaliauskiene, G., Kaminskiene, L., & Anusiene, L. (2007). Reflective Practice: Assessment of Assignments in English for Specific Purposes. *Iberica*, *14*, 149–166.

Keith, M. (1998). "Validity, Facts, and Values sans Closure: Reply to Messick, Reckase, Moss, & Zimmerman", *Social Indicators Research* 45: 73–82. Kluwer Academic Publishers

Kellaghan, T., & Greaney, V. (2001). *Using assessment to improve the quality of education* (p. 98). Paris: International Institute for Educational Planning.

Kerstjens, M., & Nery, C. (2000). *Predictive validity in the IELTS test: A study of the relationship between IELTS scores and students' subsequent academic performance*. *IELTS Research Reports*. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Predictive+validity+in+the+IELTS+test#4

Khan, M. (2019). Saudi Arabia's Vision 2030. *Defence Journsl*, *35*(3), 36–43.

Khan, P., & Iqbal, M. (2012). " Over crowded classroom : a serious problem for teachers ." *Elixir International Journal*, *49*, 10162–10165.

Khosrowjerdi, M., & Kashani, Z. S. (2013). Asian top universities in six world university ranking systems. *Webology*, 10(2), 1-9.

Khoury, I. E. (2013). *Unbridled Globalization: The Transformation of Higher Education in Qatar*. The Pennsylvania State University,United States.

Kidder, L. H., Judd, C. M., Smith, E. R. & Selltiz, C. (1986) Research methods in social relations, Published for the Society for the Psychological Study of Social Issues (SPSSI), London Holt, Rinehart and Winston.

Kim, M. M., Rhoades, G., & Woodard, D. (2003). Sponsored research versus graduating students?

Intervening variables and unanticipated findings in public research universities. *Research in Higher Education* 44(1): 51--82.

King Faisal University. (2019). *King Faisal University Strategy 2020-202*4. Concise Version. Al Hofuf.

King Saud University. (2019). *Student Guide*. Riyadh.

King Saud University. (2019). *Vision's Achievement Office*. Riyadh

Kirdar, N. (2002). *Arab Education. paper presented at the Sponsored Conference on Arab Education-AMIDEAST*. Marrakech, Morocco. Retrieved from http://www.albayan.ae

Knight, J. (2003). Internationalization remodeled: Definitions, rationales, and approaches. *Journal for Studies in International Education,* 8(1), 5–31.

Knight, J. (2004). An internationalization model: Responding to new realities and challenges. In H. de Wit, I. C. Jaramillo, J. Gacel-Avila, & J. Knight (Eds.), Higher education in Latin America: *The international dimension* (pp1–38). Washington, DC: World Bank.

Kobrin, J. L., Patterson, B. F., Shaw, E. J., Mattern, K. D., & Barbuti, S. M. (2008). *Validity of the SAT for Predicting First-Year College Grade Point Average*. , New York.

Koyame-Marsh, R. O. (2017). The dichotomy between the Saudi Women's education and economic participation. *The Journal of Developing Areas*, 51(1), 431-441.

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2001). A comprehensive meta-analysis of the predictive validity of the Graduate Record Examinations: Implications for graduate student selection and performance. *Psychological Bulletin*, 127, 162-181.

KU. (2015). *University student guide. Admission and Registration, KU*. Kuwait: Admission and Registration. https://doi.org/10.1017/CBO9781107415324.004

Kvale, S. (1996). *Interviews: An introduction to qualitative research interviewing*. Thousand Oaks, CA: Sage.

Lam, R. (2015). Studies in Higher Education Assessment as learning : examining a cycle of teaching , learning , and assessment of writing in the portfolio-based classroom. *Studies in Higher Education*, (Online publication date: 4-Apr-2016), 1–15. https://doi.org/10.1080/03075079.2014.999317

Lamadrid-Figueroa, H., Castillo-Castillo, L., Fritz-HernÁndez, J., & MagaÑa-Valladares, L. (2012). Admissions Criteria as Predictors of Students' Academic Success in Master's Degree Programs at the National Institute of Public Health of Mexico. *Public Health Reports*, *127*(6), 605–611. https://doi.org/10.1177/003335491212700612

Leonard, D. K. & Jiang, J. (1999) Gender Bias and the College Predictors of the SATs: A Cry of Despair, *Research in Higher Education*, 40, pp. 375-407.

Lesage, E., Valcke, M., & Sabbe, E. (2013). Scoring methods for multiple choice assessment in Higher Education - Is it still a matter of number right scoring or negative marking? *Studies in Educational Evaluation*, *39*(3), 188–193. https://doi.org/10.1016/j.stueduc.2013.07.001

Lin, T. J., Lee, M. H., & Tsai, C. C. (2014). The Commonalities and Dissonances Between High-School Students' and Their Science Teachers' Conceptions of Science Learning and Conceptions of Science Assessment: A Taiwanese sample study. *International Journal of Science Education*, *36*(3), 382–405. https://doi.org/10.1080/09500693.2013.780317

Lincoln, Y. S. & Guba E. G. (2000). *Naturalistic Inquiry*. New York, Sage Publications, Inc.

Linn, R. L. (1994). Fair test use: Research and policy. In M. G. Rumsey, C. B. Walker, & J. H. Harris (Eds.), *Personnel Selection and Classification*, (pp. 363–375). Hillsdale, NJ: Lawrence Erlbaum.

Linn, R. L. (2000). Assessments and Accountability. *Educational Researcher*, *29*(2), 4–16. https://doi.org/10.3102/0013189X029002004

Liu, S. (2007). *Design and use of assessment tasks in online graduate courses: Instructors' practices, reflections, and perceptions*. Indiana University.

Lockyer, S. 2004. Heard the one about…Applying mixed methods in humour research? *International Journal of Social Research Methodolog*y, 9(1): 41-59.

Lorna, E., & Katz, S. (2006). *Rethinking classroom assessment with purpose in mind. Assessment for learning, assessment as learning, assessment of learning.* Winnipeg, Manitoba: Western Northern Canadian Protocol. Retrieved from http://www.wncp.ca/media/40539/rethink.pdf

Lune, H., & Berg, B. L. (2012). *Qualitative research methods for the social sciences*. Pearson.

Lustig, K. (1996). *Portfolio Assessment: A Handbook for Middle Level Teachers*. Columbus, Ohio: National Middle School Association.

Maclellan, E. (2004). How convincing is alternative assessment for use in Higher Education? *Assessment & Evaluation in Higher Education*, *29*(3), 311–321. https://doi.org/10.1080/0260293042000188267

Madaus, G. F. (1988). The distortion of teaching and testing: High-stakes testing and instruction. *Peabody Journal of Education*, *65*(3), 29–46. https://doi.org/10.1080/01619568809538611

Makransky, G., Havmose, P., Vang, M. L., Andersen, T. E., & Nielsen, T. (2017). The predictive validity of using admissions testing and multiple mini-interviews in undergraduate university admissions. *Higher Education Research & Development*, 36(5), 1003-1016.

Markham, T ; Larmer, J & Ravitz, J. (2003). *Project based learning handbook: A guide to standards-focused project based learning for middle and high school teachers* (2nd Ed.). Novato, California: Buck Institute for Education.

Marks, D. F., & Yardley, L. (Eds.). (2004). *Research methods for clinical and health psychology*. Sage

Marshall, J. M. (2002). *Learning with Technology Evidence that technology can and does support learning.* San Diego.

Mathiasen, R. E. (1984). Predicting college academic achievement: a research review. *College Student Journal*, 18, 380-386.

Mazzei, L. (2007). CHAPTER 7: Interviewing Within a Problematic of Silence. *Counterpoints*, 318, 89-111. Retrieved September 9, 2020, from http://www.jstor.org/stable/42979792

McDonald, B. (2012). Portfolio assessment: direct from the classroom. *Assessment & Evaluation in Higher Education*, *37*(3), 335–347. https://doi.org/10.1080/02602938.2010.534763

McLoughlin, C., & Luca, J. (2001). Quality in online delivery: What does it mean for assessment in e-learning environments. In *A Proceedings of the Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education (ASCILITE)* (pp. 417–426).

McMillan J. H. (2000). *Educational research: Fundamentals for the consumer* (3rd ed). Addison

Wesley Longman, New York

McNamara, T. (2001) 'Rethinking alternative assessment', *Language Testing*, 18(4), pp. 329–332. doi: 10.1177/026553220101800401.

Merriam, S. B. (1998). Q*ualitative research and case study applications in education*. San Francisco: Jossey-Bass.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), E*ducational measuremen*t (3rd ed., pp. 13-103). New York: Macmillan.

Messick, S. (1994). *Validity of Psychological Assessment: Validation of Inferences from Persons' Responses and Performancesas Scientific Inquiry into Score Meaning. Educational Testing Service,*.

Metz, S. (2015). Project-Based Science Learning. *The Science Teacher*, *82*(1), 6.

Mikecz, R. (2012). Interviewing Elites: Addressing Methodological Issues. *Qualitative Inquiry*, *18*(6), 482–493. https://doi.org/10.1177/1077800412442818

Miles, M. B., & Huberman, A. M. (1994). Q*ualitative data analysis: An expanded sourcebook* (2nd edition). sage.

Ministry of Education. (1998). *Approved Student assessment Regulations*. Riyadh

Ministry of Education. (2004). *Education policy in the Kingdom of Saudi Arabia*. Riyadh

Ministry of Education. (2007). *Education policy in the Kingdom of Saudi Arabia*. Riyadh

Ministry of Education. (2010). *Education policy in the Kingdom of Saudi Arabia*. Riyadh

Ministry of Education. (2014). *Regulation Student Assessment*. Riyadh

Moghal, S., Kazi, A., & Bukhari, A. (2019). Large Classes and English Language Teaching and Learning in Public Sector Secondary Schools of Pakistan. *Indonesian TESOL Journal*, *1*(1), 1–8. https://doi.org/10.24256/itj.v1i1.551

Molesworth, M., Scullion, R. and Nixon, E. (2011). *The Marketisation of Higher Education and the Student as Consumer*. London, SAGE Publications: Routledge. https://doi.org/10.4324/9780203842829

Morales, M. C., & Saenz, R. (2007). Correlates of Mexican American Students' Standardized Test Scores: An Integrated Model Approach. *Hispanic Journal of Behavioral Sciences*, *29*(3), 349–365. https://doi.org/10.1177/0739986307302176

Morales-Vives, F., Camps, E., & Dueñas, J. M. (2020). Predicting academic achievement in adolescents: The role of maturity, intelligence and personality. *Psicothema*, 32(1), 84-91.

Morgan, R. (1990). Analyses of the predictive validity of the SAT and high school grades from 1976 to 1985. *Predicting College Grades: An Analysis of Institutional Trends over Two Decades.*, (89), 103–115. Retrieved from http://www.redibw.de/db/ebsco.php/search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=1991-97432-004&site=ehost-live

Moss, P. A. (1994). Can There Be Validity without Reliability. *Educational Research Journal*, 23 (2), 5-12.

Moss, P. A. (2007). Reconstructing Validity. *Educational Researcher*, *36*(8), 470–476. https://doi.org/10.2307/4621102

Murshid, K. R. (2013). The predictive value of individual admission criteria on academic performance in a Saudi medical college. *Journal of Taibah University Medical Sciences*, *8*(1), 18–23. https://doi.org/10.1016/j.jtumed.2013.01.005

Musa, M.(2003).'General aptitude tests between acceptance and rejection', *Alyaum*, (*10850)*, p. 3.

Nachmias, C. F. & Nachmias, D. (1996). *Research Methods in the Social Sciences*, 5th Edition. London: Arnold.

Naseem, S., & Dhruva, K. (2017). Issues and Challenges of Saudi Female Labor Force and the Role of Vision 2030: A Working Paper. *International Journal of Economics and Financial Issues*, *7*(4), 23–27. Retrieved from https://www.econjournals.com/index.php/ijefi/article/view/4739

NCAAA. (2010). *National Commission for Academic Accreditation & Assessment*. Riyadh, Saudi Arabia.

Nicolescu, L. (2009). Applying Marketing to Higher Education: Scope and Limits. *Management & Marketing*. 4(2): 35-44

Nolan, L. E. (2011). *Keeping the kingdom: the politics of higher education reform in Saudi Arabia.* PhD thesis, Fletcher School of Law and Diplomacy. Tufts University

Nunan, D. (1999). *Research methods in language learning*. Eighth printing. Cambridge: CUP.

Ogar, O. (2013). Assessment as a Veritable Tool for Effective Teaching and Learning. *International Letters of Social and Humanistic Sciences*, *19*, 28–33. https://doi.org/10.18052/www.scipress.com/ILSHS.19.28

Oliver, E. (2015). *Alternative Assessment For Effective Open Distance Education*. Master thesis. University Of South Africa.

Opre, D. (2015). Teachers' Conceptions of Assessment. *Procedia - Social and Behavioral Sciences*, *209*(3–4), 229–233. https://doi.org/10.1016/j.sbspro.2015.11.222

Oranye, N. O. (2016). The Validity of Standardized Interviews Used for University Admission Into Health Professional Programs: A Rasch Analysis. *SAGE Open*, *6*(3), 1–10. https://doi.org/10.1177/2158244016659119

Owen, L. (2016). The Impact of Feedback as Formative Assessment on Student Performance. *International Journal of Teaching and Learning in Higher Education*, *28*(2), 168–175. Retrieved from http://www.isetl.org/ijtlhe/

Pagano, N. . P. (2007). *Perceptions of the assessment mandate in Higher Education*. Loyola University Chicago.

Panczyk, M., & Gotlib, J. (2015). Assessment of reliability, sensitivity, objectivity and validity of MCQ Pharmacology Exams as a potential output variable for predictive analysis. *Indian Journal of Pharmaceutical Education and Research*, *49*(1), 1–9. https://doi.org/10.5530/ijper.49.1.1

Pearce, G., & Lee, G. (2009). Viva Voce (Oral Examination) as an Assessment Method. *Journal of Marketing Education*, *31*(2), 120–130. https://doi.org/10.1177/0273475309334050

Peers, I., & Johnston, M. (1994). Influence of learning context on the relationship between A-level attainment and final degree performance: a meta-analytic review. *British Journal of Educational …*, 7–132. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.2044-8279.1994.tb01081.x/full

Phelps, R. P. (2008). century have compared different teachers' evaluations of identical student work or compared the consistency of teachers' marks to those of standardized test results over time. Not surprisingly, *Nonpartisan Education Review* / ESSAYS, Vol.4, No.3, 2008. In *the World Association for Educational Research* (Vol. 4, pp. 1–9). Cadi Ayyad University, Marrakesh, Morocco. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=D13CACE6E10530861B76BFA84D 272A39?doi=10.1.1.694.6504&rep=rep1&type=pdf

Pishghadam, R., Adamson, B., Sadafian, S. S., & Kan, F. L. F. (2014). Conceptions of assessment and teacher burnout. *Assessment in Education: Principles, Policy and Practice*, *21*(1), 34–51. https://doi.org/10.1080/0969594X.2013.817382

Porta, D., & Keating, M. (2008). *Approaches and Methodologies in the Social Sciences A Pluralist Perspective* (1st ed.). New York: Cambridge University Press, New York. https://doi.org/http://dx.doi.org/10.1017/CBO9780511801938.003

Powers, D. E. (2001). Validity of GRE General Test Scores for Admission to Colleges of Veterinary Medicine. *ETS Research Report Serie*s, 2001(1), i-41.

Pring, R. (2000). *Philosophy of Education: Aims, Theory, Common Sense and Research*. Continuum Intl Pub Group

Quansah, F. (2018). Traditional or Performance Assessment: What is the Right Way to Assessing Learners? *Research on Humanities and Social Sciences*, *8*(1), 21–24. Retrieved from www.iiste.org

Rahal, A. (1987). *The ability of admission tests which are used in the Faculty of Physical Education in the prediction of the collection of students in science subjects.* University of Jordan, Amman.

Railsback, J. (2002). *Project-Based Instruction : Creating Excitement for Learning (P29)*. Northwest Regional Educational Laboratory. Portland: Published quarterly for NWREL member institutions.

Rajkhan, S. F. (2014). *Women in Saudi Arabia Status , Rights , and Limitations*. University of Washington Bothell School. Retrieved from https://digital.lib.washington.edu/researchworks/bitstream/handle/1773/25576/Rajkhan - Capstone.pdf?sequence=1

Ramist, L., Lewis, C., & McCamley-Jenkins, L. (2001). *The RedesUsing Achievement Tests SAT II Subject Tests to Demonstrate Achievement and Predict College Grades Sex Language Ethnic and Parental Education Groups*. New York: College Entrance Examination Board.

Reeves, T. C. (2000). Alternative Assessment Approaches for Online Learning Environments in Higher Education. *Journal of Educational Computing Research*, *23*(1), 101–111. https://doi.org/10.2190/GYMQ-78FA-WMTX-J06C

Reinholz, D. (2015). The assessment cycle: a model for learning through peer assessment. *Assessment & Evaluation in Higher Education*, *2938*(January), 1–15. https://doi.org/10.1080/02602938.2015.1008982

Reynolds, J. M., Blaxter, L., Hughes, C., & Tight, M. (2010). *How to research* (4th ed.). Berkshire:UK: Open University Press. https://doi.org/10.2307/1319404

Richards, L. (2009). *Handling qualitative data: A practical guide*, London: Sage.

Ricketts, C. R. (2010). *End of course grade and standardized test scores: Are grade predictive of*

student achievement? Capella University.

Riessman, C. K. (1993). *Narrative Analysis*. Newbury Park, CA: Sage.

Ritchie, J & Lewis, J (2003), *Qualitative Research Practice: A Guide for Social Science Students and Researchers'*. London: SAGE.

Rothstein, J. (2005). SAT scores, high schools, and collegiate performance predictions. *Unpublished Paper*, (609). Retrieved from http://gsppi.berkeley.edu/faculty/jrothstein/workingpapers/rothstein_cbvolume.pdf

Rugh, W. (2002). Arab education tradition growth and reform. *The Middle East Journal, 56*(3), 396–414. https://doi.org/10.1007/s10833-013-9225-6

Rumsey, L. (2013). *Shedding Light on the Predictive Validity of English Proficiency Tests in Predicting Academic Success*. University of Newcastle upon Tyne.

Sadler, D. R. (1998). Formative Assessment: revisiting the territory. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 77–84. https://doi.org/10.1080/0969595980050104

Sadler, R. (2005). Interpretations of criteria-based assessment and grading in Higher Education. *Assessment & Evaluation in Higher Education*, *30*(2), 175–194. https://doi.org/10.1080/0260293042000264262

Safón, V. (2013). What do global university rankings really measure? The search for the X factor and the X entity. *Scientometrics*, 97(2), 223-244.

Salkind, N. J. (2013). *Statistics for People Who (Think They) Hate Statistics.* (5th edition). Sage Publications, California.

Sandford, B. A., & Hsu, C.-C. (2013). Alternative Assessment and Portfolios: Review, Reconsider, and Revitalize. *International Journal of Social Science Studies*, *1*(1), 215–221. https://doi.org/10.11114/ijsss.v1i1.69

Service, E. T. (2015). *TOEFLTest and Score Data Summary*. Princeton, NJ, USA.

Shannon, J. (2008). *Reading Results: A Critical Look at Standardized Testing and the Linguistic Minority. Masters Theses. School of Education Dominican University of California*. Dominican University of California. https://doi.org/10.1017/CBO9781107415324.004

Shaw, E., Kobrin, J., Patterson, B., & Mattern, K. (2012). *The validity of the SAT for predicting cumulative grade point average by college major*. New York, NY: College Board. Retrieved from http://research.collegeboard.org/sites/default/files/info2go/2012/8/infotogo-2012-6-validity-sat-predicting-cumulative-gpa-major.pdf

Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational researcher*, 29(7), 4-14.

Simonson, M., Smaldino, S., Albright, M., & Zvacek, S. (2015). *Teaching and learning at a distance: Foundations of distance education* (sixth edit). Charlotte, U.S.A: Information Age Publishing. Retrieved from http://learning.fon.edu.mk/knigi/teachinganlearningatadistance-4.pdf

Sireci, S. G., & Talento-miller, E. (2006). Evaluating the Predictive Validity of Graduate Management Admission Test Scores, *Educational and Psychological Measurement*, 66(2), 305-317.

Smeyers, P. (2001). Qualitative versus quantitative research design: A plea for paradigmatic tolerance in educational research. *Journal of philosophy of education*, 35(3), 477-495.

Smith, L:Abouammoh, A. (2013). *Higher Education in Saudi Arabia*. (L. Smith & A. Abouammoh, Eds.) (Vol. 40). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-6321-0

Smith, M. L., & Fey, P. (2000). Validity and accountability in high-stakes testing. *Journal of Teacher Education*, *51*(5), 334–344. https://doi.org/10.1177/0022487100051005002

Snyder, V., Hackett, R. K., Stewart, M., & Smith, D. (2003). Predicting Academic Performance and Retention of Private University Freshmen in Need of Developmental Education. *Research & Teaching in Developmental Education*, *19*(2), 17–28. Retrieved from http://ezproxy.umsl.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ666647&site=ehost-live&scope=site

Soanah, A., Shalood, F., Alageal, M., & Algarni, S. (2003) *Styles and Implementations of Educational Training (second part) Ministry of Education in Saudi Arabia*: Educational Training Department

Sofi, L. A. (2015). *Teaching English in Saudi Arabia Through the Use of Multimedia*. University of San Francisco Teaching. https://doi.org/Master's Projects

Spencer-oatey, H & Dauber, D. (2019). *Internationalisation and student diversity : how far are the opportunity benefits being perceived and exploited ?* https://doi.org/10.1007/s10734-019-00386-4

Spinath, B., Eckert, C., & Steinmayr, R. (2014). Gender differences in school success: What are the roles of students' intelligence, personality and motivation? *Educational Research*, *56*(2), 230–243. https://doi.org/10.1080/00131881.2014.898917

SQU. (2015). *Academic system for university studies* (4th ed.). Muscat, Oman: Admission and Registration.

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Grounded theory procedures and techniques* (2nd ed.). Thousand Oaks, CA: Sage.

Stromquist, N. P. (2007). Internationalization as a response to globalization: Radical shifts in university environments. *Higher Education*, *53*(1), 81–105. https://doi.org/10.1007/s10734-005-1975-5

Struyven, K., Dochy, F., & Janssens, S. (2005). Students' perceptions about evaluation and assessment in higher education: A review. *Assessment & Evaluation in Higher Education*, 30(4), 325-341.

Stull, J. C., Jansen Varnum, S., Ducette, J., Schiller, J., & Bernacki, M. (2011). The many faces of formative assessment. *International Journal of Teaching and Learning in Higher Education*, *23*(1), 30–39.

Suskie, L. (2009). *Assessing Student Learning: A Common Sense Guide* (2nd Editio). San Francisco, CA: Jossey-Bass.

Talento-Miller, E. (2008). Generalizability of GMAT validity to programs outside the US. *International Journal of Testing*, 8(2), 127-142.

Taras, M. (2005). Assessment-summative and formative-some theoretical reflections. *British Journal of Educational Studies*, *53*(4), 466–478. https://doi.org/10.1111/j.1467-8527.2005.00307.

Tarawneh, M. A. (2004). The effect of using the diagnostic evaluation method on learning Arabic grammar among ninth grade students in the Kasbah schools in Ma'an Governorate. *Journal*

of the College of Education.United Arab Emirates University, 21.

Tashakkori, A., & Teddlie, C. (1998). Mixed methodology: Combining qualitative and quantitative approaches. *Applied Social Research Methods Serie*s (Vol. 46). Thousand Oaks, CA: Sage.

Tashakkori, A., & Teddlie, C. (Eds.). (2003). *Handbook of mixed methods in social & behavioral research*. Thousand Oaks, CA: Sage.

Teddlie, C., & Yu, F. (2007). Mixed methods sampling: A typology with examples. *Journal of mixed methods research*, 1(1), 77-100.

Ţeican, B. (2015). Impact of Alternative Assessment in Higher Education. *European Academic Research*, *3*(8), 8932–8942.

Tesfa, M. (2013). *The validity of University Entrance Examination and High school Grade point average for predicting first year university students academic performance*. Master's thesis. University of Twente.

Therrien, A. 2020, University of Washington eliminates standardized testing for students, Daily Hiveaccessed viewed 16 September 2020,<https://dailyhive.com/seattle/university-of-washington-removing-standardized-testing>

Tomlinson, M. (2018). Conceptions of the value of Higher Education in a measured market. *The International Journal of Higher Education Research*, *75*(4), 711–727. https://doi.org/10.1007/s10734-017-0165-6

Tshabalala, T., & Ncube, A. C. (2014). The effectiveness of measurement and evaluation in Zimbabwean primary schools: Teachers and Heads' Perceptions. *International Journal of Innovation and Applied Studies*, 8(1), 141.-148.

UAEU. (2016). *Admissions policies guide UAE university*. Al Ain, UAE.

UCLES. (2013). *IELTS Information for candidates*. Cambridge, UK. Retrieved from http://www.ielts.org/pdf/Information for Candidates_2013.pdf

Unity, O., & Igbudu, U. (2015). Influence of gender on students' academic achievement in government subject in public secondary schools in Oredo Local Government area of Edo State, Nigeria. *Journal of Educational and Social Research*, 5(2), 101-106.

Vanderstoep, S. W., & Johnson, D. D. (2008). *Research methods for everyday life: Blending qualitative and quantitative approaches*. John Wiley & Sons.

Van Teijlingen, E., & Hundley, V. (2002). The importance of pilot studies. *Nursing Standard,* 16(40), 33–36. https://doi.org/10.7748/ns2002.06.16.40.33.c3214

Van Wesel, M., & Prop, A. (2008). Comparing students' perceptions of paper-based and electronic portfolios. *Canadian Journal of Learning and Technology*, *34*(3), 1–8. Retrieved from http://cjlt.csj.ualberta.ca/index.php/cjlt/article/view/505/236

Varshney, D. (2019). The Strides of the Saudi Female Workforce : Overcoming Constraints and Contradictions in Transition. *Journal of International Women's Studies*, *20*(2), 359–372.

Vinke, A. A., & Jochems, W. M. G. (1993). English proficiency and academic success in international postgraduate education. *Higher Education*, *26*(3), 275–285. https://doi.org/10.1007/BF01383487

Volante, L. (2004). Teaching to the Test Teaching to the Test: What Every Educator and Policy-marker Shold Know. *Canadian Journal of Educational Administration and Policy*, (35), 6.

Volante, L. (2006). An Alternative Vision for Large-scale Assessment in Canada. *Journal of Teaching and Learning*, *4*(1), 1–14.

Voyer, D., & Voyer, S. D. (2014). Gender differences in scholastic achievement: a meta-analysis. *Psychological bulletin*, 140(4), 1174-1204.

Vu, L., & Vu, P. (2013). Is the TOEFL score a reliable indicator of international graduate students' academic achievement in American Higher Education? *International Journal on Studies in English Language and Literature*, *1*(1), 11–19. Retrieved from http://www.arcjournals.org/pdfs/ijsell/ijsell_3.pdf

Vulperhorst, J., Lutz, C., de Kleijn, R., & van Tartwijk, J. (2018). Disentangling the predictive validity of high school grades for academic success in university. *Assessment and Evaluation in Higher Education*, *43*(3), 399–414. https://doi.org/10.1080/02602938.2017.1353586

Wainer, H., & Steinberg, L. (1992). Sex differences in performance on the mathematics section of the Scholastic Aptitude Test: A bidirectional validity study. *Harvard Educational Review*, 62(3), 323-337.

Wait, I. W., & Gressel, J. W. (2009). Relationship Between TOEFL Score and Academic Success for International Engineering Students. *Journal of Engineering Education Washington DC*, *98*(4), 389–398. https://doi.org/10.1002/j.2168-9830.2009.tb01035.x

Wang, L., Beckett, G. H., & Brown, L. (2006). Controversies of standardized assessment in school accountability reform: A critical synthesis of multidisciplinary research evidence. *Applied Measurement in Education*, *19*(4), 305–328. https://doi.org/10.1207/s15324818ame1904

Wang, W. (2013). the Validity of GRE Performance for Engineering Students Scores on Predicting Graduate. Retrieved from http://digitalcommons.unl.edu/cehsdiss/192/

Wao, J. O., Ries, R., Flood, I., Lavy, S., & Ozbek, M. E. (2016). Relationship between admission GRE scores and graduation GPA scores of construction management graduate students. *International Journal of Construction Education and Research*, 12(1), 37-53.

Weis, M., Heikamp, T., & Trommsdorff, G. (2013). Gender differences in school achievement: The role of self-regulation. *Frontiers in Psychology*, *4*(JUL), 1–10. https://doi.org/10.3389/fpsyg.2013.00442

Whalley, W. B. (2015). Evaluating student assessments: the use of optimal foraging theory. *Assessment & Evaluation in Higher Education*, *2938*(February), 1–16. https://doi.org/10.1080/02602938.2014.991909

Wideen, M. F., O'Shea, T., Pye, I., & Ivany, G. (1997). High-Stakes Testing and the Teaching of Science. *Canadian Journal of Education / Revue Canadienne de l'éducation*, *22*(4), 428. https://doi.org/10.2307/1585793

Wiliam, D. (2010). Standardized testing and school accountability. *Educational Psychologist*, *45*(2), 37–41. https://doi.org/10.1080/00461521003703060

Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, *37*(1), 3–14. https://doi.org/10.1016/j.stueduc.2011.03.001

Wilkinson, D., Zhang, J., & Parker, M. (2011). Predictive validity of the Undergraduate Medicine and Health Sciences Admission Test for medical students' academic performance. *Medical Journal of Australia*, *194*(7), 341–344. Retrieved from https://www.mja.com.au/journal/2011/194/7/predictive-validity-undergraduate-medicine-and-health-sciences-admission-test

Yan, Z. (1995). *Predictive validity of TOEFL scores on first term's GPA as the criterion for international exchange students*. PhD Thesis. The University of British Columbia. Retrieved from http://medcontent.metapress.com/index/A65RM03P4874243N.pdf

Yang, R. (2003). Internationalization: concepts, complexities and challenges. *International Review of Education*, *49*(3–4), 269–291. https://doi.org/10.1023/A

Yanosky, R. (2009). Institutional Data Management in Higher Education. ECAR Key Findings. *Educause*

Yeain, O. P. (2003). *The Predictive Validity Of Başkent University Proficiency Exam (Beep) Through The Use of the Three-Parameter Irt Model Ability Estimates*. PhD Thesis. Middle East Technical University.

Yen, D., & Kuzma, J. (2009). Higher IELTS Score, Higher Academic Performance? The Validity of IELTS in Predicting the Academic Performance of Chinese Students. *Worcester Journal of Learning and Teaching*, *3*, 1–7. Retrieved from http://eprints.worc.ac.uk/811/1/YenKuzmaIELTScores.pdf

Yin, R. K. (2003). Case Study Research: Design and Methods, (3rd ed). *Applied Social Research Methods Series*. Thousand Oaks, CA: Sage Publications.

Young, J. W. (2001). Differential Validity, Differential Prediction, and College Admission Testing: A Comprehensive Review and Analysis. Research Report No. 2001-6. *College Entrance Examination Board*.

Young, J. W., Klieger, D., Bochenek, J., Li, C., & Cline, F. (2014). The validity of scores from the GRE revised general test for forecasting performance in business schools: Phase one. *ETS Research Report Serie*s, 2014(2), 1-10.

Zimbicki, D. (2007). *Examining the Effects of Alternative Assessment on Student Motivation and Self-Efficacy*. Walden University, Washington, US.

Zohrabi, M. (2013). Mixed methods research: Instruments, Validity, Reliability and Reporting Findings. *Theory and Practice in Language Studies*, *3*(2), 254–262. https://doi.org/10.4304/tpls.3.2.254-262

Zwick, R. (Ed.). (2013). *Rethinking the SAT: The future of standardized testing in university admissions.* Routledge.