COMPUTER ASSISTED ASSESSMENT IN OMAN: FACTORS AFFECTING STUDENT PERFORMANCE

Amina Obaid Al-Hajri

A thesis submitted to the University of Plymouth in partial fulfilment for the degree of

DOCTOR OF PHILOSOPHY

School of Computing & Mathematics
Faculty of Science & Technology

2011

COMPUTER ASSISTED ASSESSMENT IN OMAN: FACTORS AFFECTING STUDENT PERFORMANCE

Amina Obaid Al-Hajri

A thesis submitted to the University of Plymouth in partial fulfilment for the degree of

DOCTOR OF PHILOSOPHY

School of Computing & Mathematics
Faculty of Science & Technology

2011

Amina Obaid Al-Hajri

COMPUTER ASSISTED ASSESSMENT IN OMAN: FACTORS AFFECTING STUDENT PERFORMANCE

Abstract

This thesis investigates the social and psychological factors that might affect Omani higher education students if computerised assessment was to be implemented. A review of the literature and the historical and cultural development in Oman suggested that a number of different variables might affect students' performance when taking computerised assessment. These factors which include gender, college of study and geographical region of residence may cause unwanted and selective differences in student performance which are not related to the content of the assessment. In addition, the potential effects of such variables as computer experience and computer self-efficacy on student performance were investigated. The study also explored student and academic staff attitudes towards computerised assessment.

Both quantitative and qualitative methods are used in this study through a selection of instruments such as a test that was delivered in different modes, questionnaires, focus groups and semi-structured interviews. Quantitative approaches are used to provide answers to the main study questions about student performance, and qualitative approaches are used to get deeper insights into the students' and staff members' perceptions, attitudes and values in relation to the research subject of the study. All these instruments were developed based upon the literature and also validated through a separate initial study. The main study took place after the instruments had been validated and involved over 400 students and 100 staff at three Omani Applied Sciences Colleges. Statistical analysis showed a small but significant difference between the two assessment modes in favour of the paper-and-pencil test. There was a significant difference in performance between both genders, with females out-performing males. However, the most striking finding was a differential effect of assessment mode between males and females. Males performed better in the computerized test than in the paper-based one, in contrast to females whose performance in the paper test was better. This suggests that the introduction of computerised testing may affect males and females in different ways. One

possible explanation for this is that Omani males have more opportunities to use computers inside and outside homes.

The questionnaire results and the qualitative information from focus groups both showed that females were more nervous and found it more difficult to read from the computer screen than males did. Also, it was obvious in the focus group discussions that females had more negative feelings towards computerised testing compared to males, both before and after experiencing computerised assessment.

The study found that students' performance in the English language test had showed significant variation across colleges, and among students from different regions This variation seemed to be associated with variation in computer experience among students at the different colleges and from different regions This may be due to regional differences or specificities, especially in terms of computer use, among the nine administrative regions in Oman

Staff attitudes and perceptions towards CAA, in general, were positive and not affected by either the gender or nationality/language factor. Most of the academic staff members revealed their willingness to implement CAA but also stressed that CAA should be gradually implemented. Both students and staff members identified a number of important points such as the need for a reliable system, qualified technicians and sufficient computers if Computer Assisted Assessment was to achieve wide acceptability.

Table of Contents

List of Tables

List of Figures

Acknowledgements

First and foremost, I offer my sincere thanks to my supervisor, Dr Chris Ricketts. I gratefully acknowledge his enormous support and astute guidance on all aspects of this project. He has added considerably to this thesis through his intelligent comments and insightful suggestions. I am also grateful to him for his patience, encouragement and commitment to this thesis. I would also like to extend my thanks to Professor Andy Hannan to his guidance and useful advice during the first year of my study and to Dr Sam Regan de Bere for her guidance in the qualitative aspect of this thesis.

Very special thanks to Mr Mohammad Al-Hajri, from Sur College, for his technical assistance and support and for his generous and continuous help. Also I would like to thank the Officials in the Ministry of Higher Education in Oman.

I would like to extend my appreciation to Mr Mahrez Bachta for supporting me in various ways.

Special thanks to my family members, especially my mother, for her unremitting emotions, encouragement and love.

Last but not least, I would like to express my gratitude to my husband, Waleed Al-Harthy, for his loving, caring, supporting, motivating, patient encouragement, and his indefatigable help and assistance.

Finally, this thesis is lovingly dedicated to my daughter, Joud.

Author's Declaration

I declare that this thesis, submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the Department of Computing and Mathematics, Faculty of Science & Technology, University of Plymouth, is wholly my own work unless otherwise referenced or acknowledged. The document has not been submitted for qualifications at any other academic Institution.

This study was financed from the Ministry Of Higher Education, Sultanate of Oman.

I have one publication which processed to the "2010 international Computer Assisted Assessment (CAA) conference". The title of the paper that I presented at the conference is "gender differences in CAA performance in Oman: Is it part of the culture?.

Word count of main body of thesis: 60 000 word (starting from chapter one excluding references and appendices).

Signed.......................................

Date.....23/03/2010..........................

Chapter One: Introduction

1.1. Introduction

Despite the numerous steps which have been taken to improve the educational system in Oman, and the substantial advances which have been made in this sector so far, assessment is still carried out in the traditional paper-and-pencil mode. And as the use of IT has become a common practice worldwide, a new trend enhancing the use of the computer as an educational tool has started to gain popularity. This includes using computers not only in designing and delivering lessons, but also in setting tests and conducting assessment.

Hence, the importance of this research project which seeks to examine the various effects of the different modes of assessment administration, and investigate the potential factors which could affect the transition from the paper-and-pencil administration mode to the computer-based one.

As computer prices are constantly decreasing, the use of the computer both at home and in schools has witnessed a noticeable increase. This has reinforced the tendency to use computers for educational purposes, particularly in higher education institutions. In this respect, Larson (1987, p.20) points out that "the use of computers has infiltrated many areas of education [and] although sophisticated computer programs have historically been more common in a lot of countries, we are now beginning to see several worthwhile applications in the field as well".

It is worth noting that many employers, psychologists, educators and researchers have already converted the conventional or paper-based tests which they use into computer-based test formats. One example of this could be the use of computerized tests in military section and assessment in Canada, France, Germany, the Netherlands, the United Kingdom, and the United States of America (Neil, 1996).

This substantial increase in the use of computerized testing has also been paralleled by an increasing use of computers in psychological and educational assessment, primarily in the United States of America (McDonald, 2002). As cited in Smith (2003, p.2), *"this is demonstrated by the computerization of a* number of cognitive tests such as the Raven's Progressive Matrices (Waterfall, 1979), the Peabody Picture Vocabulary Test (Space, 1975), Graduate Record Exam (GRE: Mills, 1999; Schaeffer, Reese, Steffen, Mckintey and Mills, 1993) and the Test of English as a Foreign Language TOEFL (Stricker and Wilder, 2001)."

Historically, the first attempt to assess students by computer took place in 1959 when Rochester Polytechnic Institute in the USA first used a computer program to test the behaviour of students' machine language submissions (Winters, 2004) Since that time, CAA has started to be widely used in the USA, and according to Bennett (1998) one million students who were engaged in MA and BA programs were assessed through CAA under the auspices of a national testing program in the 1997-1998 academic year.

In addition, in Australia CAA is commonly and noticeably used at the University of Sydney as well as at the Curtin University of Technology where according to Sly & Rennie (1999), around 30,000 students sat for computer- aided summative assessment.

In 1995, the surveys conducted about the use of CAA in the UK revealed that this new mode of assessment was increasingly used, particularly at the universities where such subjects as science, computing, medicine, mathematics, engineering, physiology and modern languages are taught (Stephens & Mascia, 1995). Also, according to Stephens & Mascia (1997), 73% of the students in the UK higher education institutions had sat for at least one computer-based test by 1997.

Despite the increasing use of computer-based testing in some countries, only few educational institutions have actually adopted this new assessment mode and very few academics have really tried to apply it in some other countries (Davidson, 2003). Thus, paper-based tests are still today commonly used in the majority of higher education institutions worldwide (Davidson, 2003).

Before proceeding to highlight the importance of computer-based testing and weighing up its benefits and shortcomings, we should note that CAA is defined by Bull & McKenna (2004) as "a common term for the use of the computer to deliver, mark and analyse assignments or examinations" (p.7). Larson (1987) also defines CAA as the process which

> "Includes any use of the computer that aids the testing process. The type of assistance may be in a form of test-item generation, test delivery, scoring, record keeping, reporting results, providing feedback to examinees, and the like. While (CAA) programs vary considerably in the range of assistance they provide to testers, ideally they should eliminate as much of the drudgery of testing as possible" (p.20).

The idea of conducting this research project has been triggered by the fact that CAA has many advantages such as providing academic staff with timely, comprehensive and diagnostic feedback on students' achievement so that they could remedy any shortcomings in the curriculum or problems faced by the students. Other advantages of CAA consist of saving marking and score reporting times besides allowing flexible test scheduling; improving the balance of assessment methods; enabling assessors to measure response time and reducing both effort and the probability of human error as well as decreasing administrative costs. Bull & McKenna (2004) point out that "CAA provides assessment marks in electronic format and, therefore, the potential exists to make the administration and management of assessment data more efficient by automatically entering marks into student record systems and management information systems. The approach can save time and effort and reduce clerical errors" (p.7, 8).

However, CAA also has some disadvantages that will be carefully analyzed and dealt with in the coming chapters. These disadvantages include some validity-related issues in addition to the risks that might be associated with using technology; and more importantly the money and time needed for designing, conducting and invigilating computer-based tests; and definitely the need to ensure the smooth transition to this new assessment mode, for this shift would make an important change in assessment culture (Bull, 1999a; Davidson, 2003, Goldberg & Pedulla, 2002; Larson, 1987; Ricketts & Wilks, 2002a; Sheader, Gouldsborough & Grady, 2006; Spray et al., 1989; Wang, 2004).

In fact, every assessment type has its advantages and disadvantages, so what makes educationalists prefer one particular type to the others is the fact that the chosen type may have more advantages. Hence, the introduction of a new assessment type should be justified by its having more advantages than the others (or the traditional methods) otherwise it would be doomed to failure and rejection.

Current student assessment in the Omani educational system is based on the traditional paper-and-pencil testing mode. So, this research project will try to examine and compare the effect of the different modes of test administration and the factors which might affect students' performance in Oman. Although there are many studies which have focused on the comparison between the computer-based and the conventional paper-based testing modes, the present study differs in the following important respects:

1. Computer use in Arab countries was an uncommon practice in the past, but has rapidly increased in recent years. This study will try to explore the possibility of introducing computer-assisted testing in a culture with a short history of computer use.
2. The assessment culture is still traditional in Arab countries. This study aims to examine the possibility of introducing Computer-Assisted Assessment in a culture where both academics and students are used to traditional assessment modes solely.
3. The Sultanate of Oman is unusual in its having very few large cities and is also divided into various administrative areas which substantially differ in terms of computer and internet availability and use, both now and in the past. However, Omani students often need to attend colleges which may be situated in other areas. Hence, this study aims to discuss and assess the way in which differences in culture and computer experience might impact on students' performance in CAA, and their reaction or attitudes towards it.

It is hoped that this research opens new theoretical and practical gates for Arab researchers and enable institutions in Arab countries to make use of technological developments, and particularly of computers in introducing and applying new assessment methods. No doubt, the introduction of these methods to improve measurement or assessment tools will help to serve these fields and contribute to the development of the measurement process. This, in

turn, will make the processes related to diagnosis orientation and decision-
making more accurate and precise.

This study also aims to contribute to the current studies conducted worldwide in this area besides raising issues that may be of great interest to researchers in this field.

When intending to apply CAA, score equivalence between computer-based and the conventional paper-based tests has to be established through measuring and examining the qualitative and quantitative analysis of the test construct and the psychometric properties of the tests' forms (Staples & Luzzo, 1999, Sawaki, 2001, p.5). These steps will help to ensure the test is fair and the students are *not disadvantaged in any of the modes they would choose Peak (2005) notes* that "the transition from paper-and-pencil to computerized tests cannot be taken for granted" (p.1). Moreover, the test takers should respond to the item content rather than the way in which it is presented (Pommerich & Burden, 2000).

In this research project, I will try to address test equivalence from both quantitative and qualitative perspectives so that we could provide a conceptual basis for the study of test equivalence (Mead & Drasgow, 1993).The quantitative dimension is concerned with demonstrating numeric score comparability or the extent to which both assessment modes (computer-based and paper-based) yield comparable scores. This will also be demonstrated through the examination of psychometric criteria. The American Psychological Association (APA) considers that equivalence could be achieved if the means, dispersions and shapes of the score distributions are approximately the same (American Psychological Association, 1986) The qualitative dimension will focus on exploring students' and academic staff's perceptions, views, experience and thoughts about CAA.

Finally, this thesis is aimed to investigate the possibility of applying CAA in Omani higher education institutions, focusing on the potential variables that might impact on such application and the difficulties and potential obstacles that CAA might face from students' and academic staff's points of view.

1.2. Overview of Thesis

This study is organized in the following way:

Chapter two will be devoted to talking about assessment from a theoretical point of view by discussing its definition, properties of good assessment and how it affects student learning. Other aspects about CAA will be presented and discussed in this chapter as well. These details will include the benefits (and also shortcomings) of CAA, and the utility of assessment.

In chapter three, focus will be laid on the Omani culture and its impact on education. I will also try to review the progress and development which have been witnessed by education and higher education in Oman, as well as the history of computer use in the country.

Many of the studies which have compared students' performance on computer-based and paper-based tests and addressed the impact of the demographic factors on students' performance when taking CAA will be presented and discussed in chapter four.

Chapter five will focus on the thesis questions as well as its hypotheses. Also, it will describe the methodology, the subjects and the samples of this research. In addition, this chapter will explain how the research is designed along with its variables, the instruments that are going to be used and their validity and reliability as well as the ways of collecting and analyzing data.

Chapters six, seven, eight, and nine will form the core of the thesis and will address the results provide a summary of the findings in light of the literature.

Finally, a discussion of the findings and a conclusion will be presented in chapter ten to summarise all the study results and findings.

Chapter Two: Assessment

2.1. Introduction

"Assessment is one of the most powerful drivers of innovation and change in education, as it defines the goals for both learners and teachers" (Clark, 2003, p.32).

Assessment is a normal natural process which has happened ever since humanity started to exist (Al-Hajri, 2005). It has always taken many shapes and ways and used different tools and instruments. Humans have always assessed people's actions, behaviour and transactions. They have always been involved in assessment while bringing up and teaching their children or anyone who is in a learning position, consciously or unconsciously, voluntarily or involuntarily, in an attempt to assess and judge whether or not they have attained the required skills. These knowledges and skills are not only stored and perceived as new experiences or schemata, but passed on to new generations for development and progress, and considered as an educational repertoire (Al-Hajri, 2005).

In the educational field, assessment is not easy to define as it has always been a controversial notion. Collins Dictionary (2001), for example, defines it as "the act of assessing, especially in Britain, the evaluation of a student's achievement on a course". A similar definition is provided by Longman Dictionary of Applied Linguistics (1985) in which assessment is defined as "the measurement of the ability of a person or the quality or success of a teaching course" (Richards, Platt & Weber, 1985, p.18). Another definition for assessment is provided by Longman Dictionary of Language Teaching and Applied Linguistics (3rd edition) according to which assessment is perceived as "a systematic approach to collecting information and making inferences about the ability of a student or the quality or success of a teaching course on the basis of various sources of evidence"(Richards & Schmidt, 2002, p.35).

It is obvious, however, that these definitions do not cover all the aspects associated with this notion, nor do they address its complexity, for "assessing students' progress and performance is a complex process that involves many different elements [as] it encompasses assessment purposes and principles, content and methods, criteria and standards" (The Quality Assurance Agency for Higher Education, 2007, p.1).

As reflected by these definitions, assessment cannot be defined without addressing its aims which should not be considered as a mere component of the process of assessment but rather central to it. Angelo (1995, p 11) points to those ultimate aims in defining assessment which he considers to be "a means for focusing our collective attention, examining our assumptions, and creating a shared culture dedicated to continuously improving the quality of higher learning". Angelo's definition highlights the purposeful nature of assessment which should also be perceived as a means that serves to enable us to examine, revise and improve the teaching/learning process. Hence, assessment should be carried out for diagnostic and developmental purposes as it enables us to generate input to direct future instructional attention

The same view is also reflected by O'Farrell's (2004, p.22) definition in which she defines assessment as "the systematic and ongoing method of gathering, analysing and using information from measured outcomes to improve student learning in terms of knowledge acquired, understanding developed and skills and competencies gained". It can be concluded that both definitions view assessment as a continuous process of data collection which takes place over time and aims to examine, analyse and assess learners' achievement in order to improve their learning, rather than a process which is meant to give a one-shot judgmental act evaluating students' achievement. Hence, the need to highlight the developmental dimension of assessment. That is, repeated assessment practice promotes the better retention of course content, which results in improved learning (testing-effect) as asserted by cognitive psychology studies carried out by Larsen, Butler & Roediger (2008). This view is also reflected by Brown (2004, p.81) who concludes that "assessment is perhaps the most important thing we can do to help our students to learn".

A more comprehensive definition would be Erwin's in which assessment is perceived as "the process of defining, selecting, designing, collecting, analysing, interpreting and using information to increase students' learning and development" (Erwin, 1991, p.15).

This last definition points to the various components or processes making up assessment and their substantial role in improving students' learning by improving teaching practices in the light of the findings generated by the feedback instructors/assessors get and also by reshaping "institutional policies, processes, and practice in ways that lead to improving [.......] institutional functioning" (Peterson et al., 1999, p.4).

At a time when computers are widely used to support learning, one could but reflect on how best we can use them to improve assessment, and enhance learning.

2.2. Properties of Good Assessment

Assessing students is a complex process that involves the interplay of many factors that should be taken into consideration while devising an assessment instrument. There are many definitions of what makes good assessment. For example, according to Brown, Race & Smith (1996), assessment should have the qualities illustrated in the table 1 below which are cited in Tsintsifas (2002, p 19).

Assessment must be	**In order to:**
Valid	Accurately assess the delivered material
Reliable	Promote consistency between assessors
Fair	Offer fair opportunity for success
Equitable	Be indiscriminating between students
Formative	Give many opportunities to learn through feedback
Well timed	Provide learning stimulus and be fair
Incremental	Increase reliability and consistency over a period of time
Redeemable	Allow a series of opportunities
Demanding	Challenge students and ensure high standards
Efficient	Be manageable within the constraints of resources

Table 1. Summary of Qualities of Assessment

Also, as QAA stresses that good assessment cannot be attained unless high-quality assessment is encouraged (QAA, 2006). So, a set of recommendations and suggestions are made to encourage the adoption and implementation of high- quality assessment methods including such values as validity, reliability, consistency, transparency, fairness and formative feedback, which conforms with Brown, Race & Smith's recommendations presented in table 1 above.

Brown, Bull & Pendlebury (1997) argued that changing assessment procedures is not an easy process, for it is accompanied by several risks that should be taken into consideration and cleared before opting for any change. When the process of changing any assessment procedure is considered or approached such parties as "opinion, leaders, stakeholders, external forces and equilibria" have to be taken into account (Brown, Bull & Pendlebury, 1997, p.222, 223).

One possible approach is to consider the 'utility' framework of Van Der Vleuten (1996). This is determined by various values or variables like validity, reliability, educational impact, acceptability and cost. According to Van Der Vleuten (1996, p.55) the utility of assessment is defined as " a multiplicative function of these variables". That is, according to Van Der Vleuten, the utility of an assessment procedure is determined by the following formula:

$$U = V \times R \times E \times A \times {}^{1}/_{C}.$$

U: Utility.

V: Validity.

R: Reliability.

E: Educational Impact (impact on Learning).

A: Acceptability.

C: Cost.

This means that the utility of any assessment method is judged by the combination of its validity, reliability, acceptability, educational impact and cost. These five aspects may be perceived in different ways. "The weights of the criteria depended on how the importance of each of the different criteria was perceived by those responsible for assessment in a certain assessment situation or assessment context" (Van Der Vleuten & Schuwirth, 2005, p.309).

That is, for an assessment procedure to be useful, it needs to be sufficiently valid and reliable, accepted by stakeholders as well as students and academics besides having a positive impact on learning and also a manageable cost. On the contrary, "if one of the elements is zero, the utility will be zero. A reliable, valid and feasible test will have a short life if it's accepted by no one" (Van Der Vleuten, 1996, p.55).

The traditional perception of validity and reliability could alter when applying CAA as the latter allows for the possibility of administering the same measurement instruments more than once (test-retest), elevating students' progress levels with every administration. This means that it is unlikely or even impossible to get the same results when using the same test again. "Basically, the message is that no method is inherently unreliable and any method can be sufficiently reliable, provided sampling is appropriate across conditions of measurement" (Van Der Vleuten & Schuwirth, 2005, p.312).

The notion of the educational impact or effect of assessment on learning has been gaining large acceptance as research has proven that there is a significant correlation between the three following variables: assessment, teaching and learning. The idea of assessment driving the learning process has been widely acknowledged, which has enhanced a substantially increasing tendency for adopting and applying new assessment methods Nevertheless, this does not mean that it is easy to change traditional assessment methods as there are still many questions that need to be answered and issues to be settled and agreed upon Such questions include: how to link Intended Learning Outcomes (ILOs) with assessment, how to provide and / or increase the frequency of providing students with formative assessment and adequate feedback, how to balance formative and summative assessment, how to determine the frequency of assessment and its distribution throughout a course and many such questions.

2.2.1. Validity

"Validity refers to the degree to which evidence and theory support the interpretations of test score entailed by proposed uses of test" (American

Educational Research Association, American' Psychological Association & National Council on Measurement in Education, 1999, p.9).

Another definition is provided by Messick (1989, p.13) who argues that validity is " an integrated judgement of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores and other modes of assessment".

The process of establishing validity requires gathering evidence to provide a scientific foundation for the result interpretations in the light of the purpose of assessing so the American Educational Research Association, American Psychological Association & National Council on Measurement in Education (1999) identified various sources of validity evidence and they emphasised that validity is a unitary concept. Moreover, the traditional facets of validity such as content validity; construct validity, concurrent validity and predictive validity, have been replaced by new terms such as evidence based on test content, evidence based on response process, evidence based on internal structure and evidence based on relation to other variables.

Evidence based on test content refers to the "themes, wording, and format of the items, tasks or questions on a test" (Ibid). The evidence can be accumulated from expert judgment which aims to determine how much an assessment instrument (test) adequately and sufficiently measures the skills it sets out to measure (Carmines & Zeller, 1979). Evidence can also come from the "empirical analysis of the adequacy with which the test content represents the content domain" (American Educational Research Association, American' Psychological Association & National Council on Measurement in Education, 1999, p.11). In the field of second or foreign language teaching, for example, a test which is meant to assess communicative competence would have low evidence based on test content if it tested only some of the knowledge of the grammar and vocabulary of the target language as communicative competence also includes knowledge of the rules of speaking (such as how to begin and end a conversation, types of speech events, address forms) knowledge of using and responding to different speech acts (such as requesting, apologising, thanking), and knowledge of how to use language appropriately in a given social context.

Evidence based on response process come from analysing individuals' responses through asking test takers about their performance towards particular items which will show evidences that will enhance the construct definitions Evidence can also be gathered through analysing the relationship among the test parts and the whole test as well as other variables. All this evidence will help in interpreting the differences in the test score among the subgroup of test takers which "can assist in determining the extent to which capabilities irrelevant or ancillary to the construct may be differentially influencing their performance" (American Educational Research Association, American' Psychological Association & National Council on Measurement in Education, 1999, p.12).

Other evidence can be accumulated by analysing the internal structure where the analysis can reveal the relationship between the test items and the test components. The evidence gathered can reveal one component or may reveal several components "that are each expected to be homogeneous, but that are also distinct from each other" (Ibid, p.12).

There is also evidence based on relations to other variables external to the test which will provide important sources of validity evidence and support the interpretation of the result (Ibid, 1999). So, validity tells us whether or not the results reflect what they are supposed to measure and are not affected by another factor. According to Biggs (2003), "in the measurement model, the test needs to be validated against some external criteria to show that the trait is being measured" (p. 164).

Finally no method is inherently valid or invalid as validity relates to the interpretation of the test scores, and inferences made from them. Recently it has been argued that the modern concept of validity consider all aspects of assessment quality (William, 2008).

2.2.2. Reliability

Reliability relates to the degree of consistency of measurement. A test is said to be reliable only if it gives the same results when it is given on different occasions or when it is marked by different assessors while still targeting the same or similar cohort of students. O'Farrell (2004, p.24) notes that "reliable measures are measures that produce consistent responses over time ". However, different techniques have been developed to estimate the reliability of an assessment instrument. Reliability is usually estimated in two ways. The first way is test-retest in which the student should get the same score when the test is administered on different occasions. This involves two administrations of the same measurement instrument. The second way is the internal consistency (Cronbach's alpha and split half). Cronbach's alpha splits the test questions or items in order to measure the degree to which the items are homogeneous or consistent with each other.

Also Biggs (2003, p.163) distinguishes between two types of reliability, intra-and inter-judge reliability. Intra-judge reliability is judged by whether the same assessor makes the same judgment about the same performance on two different occasions, like the test-retest process While Inter-judge reliability is judged by whether different assessors make the same judgment about the same performance on the same occasion.

The notions of reliability and consistency are given substantial emphasis as indispensible values that should underpin high-quality assessment, for the latter should be based on reliable measures which always produce consistent results. Hence, higher education institutions are urged to publicise and use "clear assessment criteria and, where appropriate, marking schemes, are key factors in assuring that marking is carried out fairly and consistently across all subject" (QAA, 2006, precept 7, p.16).

2.2.3. Aims of Assessment (Educational impact)

It seems that whenever the word assessment is used, such words as tests, quizzes, exams spring to mind. But have we ever reflected on the reasons for conducting assessment? Brown (2004), states that the main reason for conducting assessment is to help students to learn. This general view about the main aim of assessment needs to be thoroughly analysed so that we can realise the importance of assessment and the role it plays in enhancing students' learning.

There are numerous references in the Code of Practice for the assessment of students in UK (Section 6) to the major aims of assessment in higher education institutions which are mainly: maintaining academic standards and encouraging effective learning (QAA, 2006). However, we should note that both aims are closely interrelated and represent two faces of the same coin. Maintaining academic standards can never be achieved in isolation of encouraging effective learning and similarly encouraging effective learning would lead to maintaining academic standards.

Harris & Bell (1990); and Brown, Bull & Pendlebury (1997) argue that assessment is carried out for the following reasons which ultimately serve to enhance and improve students' learning:

- Evaluating learners' performances in order to improve the quality of teaching as well as the quality of curriculum design.
- Providing students with feedback on their performances and achievements.
- Motivating students, for assessment could be viewed as a powerful extrinsically motivating tool. But for assessment, students are unlikely to engage actively in learning altogether.
- Categorising students.
- Keeping a record of students' progress

2.2.4. Assessment for Learning

It is hard to deny the educational impact of assessment and the fact that assessment is crucial for learning, for its *raison d'être* is to improve teaching as well as learning. This view is frequently stated in the literature; Black & Wiliam (1998), for example note that "all those activities undertaken by teachers, and their students in assessing themselves, [......] provide information to be used as feedback to modify the teaching and learning activities in which they are engaged "(p.143). Therefore, the success of any teaching and learning practice greatly depends on the way assessment is perceived and carried out. Rowntree asserts that "if we wish to discover the truth about an educational system we must first look to its assessment procedures" (Rowntree, 1987, p.1).

Consequently, any attempt meant to improve student's learning should include improving assessment methods and techniques This view is also strongly advocated by Brown, Bull & Pendlebury (1997, p.6) who note that " *if you want to change student learning, then change the methods of assessment*".

Assessment can not only be viewed as a stimulus for learning but could rather be regarded as an act of learning in itself. Wood writes that "assessment not only drives learning, [but] it may also help learning" (2009, p.5).

2.2.5. Feedback

If we assume that one possible educational impact of assessment is to help students to learn, then feedback must be given a prominent position in the assessment process, for it is mainly through feedback that instructors / assessors and students improve their performances. This view is highlighted in the literature and advocated in most of (if not all) the studies conducted on this topic. Brown (2004), for example, stresses the fact that generating and acting upon feedback should be set as a condition for learning. But for feedback, assessment cannot be claimed to be constructive or contributing to learning.

However, precept 19 (QAA, 2006, p 20) stresses that providing students with feedback should "not increase the burden of assessment". This could, in fact, be viewed as somehow contradictory as it is obvious that "formative feedback [...] needs to be detailed, comprehensive, meaningful to the individual, fair, challenging and supportive, which is a tough task for busy academics" (Brown, 2004, p.85) Hence, this might trigger the idea of looking for other ways which could cater for providing students with formative timely feedback without being considered over-burdensome by busy academics.

In fact, the importance of feedback stems from its tendency to provide instructors / assessors with the opportunity to carry out an ongoing process of reflection on and assessment of their performances before acting accordingly. Although feedback is formative for teachers, it is not always so for learners unless it is timely, sufficiently detailed, meaningful, regular and learning-oriented as represented by the diagram below (Figure 1).

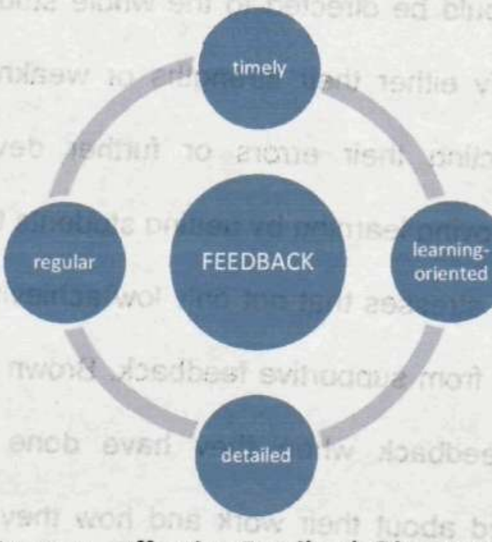


Figure 1. Effective Feedback Diagram

It is crucial that feedback is provided frequently enough and at the right time so that it could be useful to students. Cited in Bull & McKenna (2004, p.5), Falchikov (1995) and Schmidt et al., (1990) argue that "feedback needs to be accurate and constructive, and regular, formative feedback has been shown to have a marked improvement on students' overall performance on a course". Nevertheless, given too soon, feedback might prevent students from reflecting on and assessing their own work. This view is pointed out by O'Farrell (2004, p.14) who stresses that feedback "should not be provided too soon as it could prevent students from reflecting on their work; neither should it be provided too late when it is no longer salient to the student".

Feedback should also be explicit and understandable in order to be constructive so that learners could act upon it to improve their learning, for providing students with feedback would, otherwise, be pointless if they cannot figure out what is actually communicated to them. Feedback should be learning- oriented rather than focusing on marks so that it fosters learning.

However, we should stress that feedback is not meant to benefit one particular type of students, but should be directed to the whole student population as it helps students to identify either their strengths or weaknesses so that they could engage in correcting their errors or further developing their work accordingly. Hence, improving learning by getting students to get involved in the learning process. Brown stresses that not only low-achieving students but also good ones could benefit from supportive feedback. Brown points out that good students "also need feedback when they have done well to help them understand what is good about their work and how they can build on it and develop further" (Brown, 2004, p. 84). Feedback also helps to build self-confidence and boost motivation by positively reinforcing students' good works, for "students are motivated by feedback on their works" (Bull & McKenna, 2004, p.5).

However, we should not deny the fact that providing formative individual feedback is not an easy task for instructors / assessors as it is a "tough [time consuming] task for busy academics" (Brown, 2004, p.85), especially with the current rapidly-growing number of students enrolled in colleges and universities. Thus, we need to develop new assessment techniques which can cater for providing formative individual feedback for students without adding to the heavy burden which is already carried by academics. This may lead us to consider the possibility of adopting Computer Assisted Assessment as a potential alternative to traditional paper-based assessment methods.

2.2.6. Test Enhanced Learning

Larsen, Butler & Roediger (2008), stress that assessment serves not only to enhance but also to facilitate learning. According to them, the continuous and repetitive assessment of students could have two types of effects, direct and indirect. The indirect effect relates to the role of assessment in driving students to learn, the direct effect relates to the fact that the more tests students take while learning, the better retention of course content they achieve. Larsen, Butler & Roediger (2008) point out that "research in cognitive psychology has shown that tests can also directly affect learning by promoting better retention of information, a phenomenon known as the testing effect" (p.959). They stress that repeated testing results in better retention of the course content than does repeated studying, particularly because the former involves the effortful retrieval and active processing of information. They conclude that "taking a test leads to better retention than re-studying the material for an equivalent amount of time" (p.961). However, they argue that for the testing effect to take place, and to ensure long-term retention of course content, assessment should be well spaced over time. Accordingly, they assert that "tests should be given often and spaced out in time to promote better retention of information" (Ibid). It is also made clear that assessment could only serve the purpose for which it is set (enhancing learning) provided that it is accompanied by providing students with constructive feedback as "the mnemonic benefits of testing are further enhanced by feedback, which helps students to correct errors and confirm correct answers" (Larsen, Butler & Roediger 2008, p.959). Both the role of assessment in enhancing learning and the role of feedback in facilitating it have become axiomatic and are often referred to in the literature (Brown, Bull & Pendlebury, 1997; Brown, 2004; Bull & McKenna, 2004; Gibbs & Simpson, 2003).

Larsen, Butler & Roediger (2008) also distinguish between two types of assessment items, multiple-choice questions (MCQs) which require students to recognise an answer and open-ended questions (OEQs) which require students to construct an answer. According to Larsen, Butler & Roediger (2008, p.961) the latter has a better impact on students' retention of information as "research has shown that production tests lead to better retention than recognition tests, presumably because production tests require more effortful retrieval of information from memory than recognition tests".

However, we should note that it is not always feasible or practical to assess students repeatedly and provide them with accurate and constructive feedback, especially as students' numbers are rapidly increasing. Besides, academics are already overloaded with various responsibilities, which prevents them from sparing the effort and time to continuously assess their students. Bull & McKenna (2004, p.5) note that "large student groups often mean that academic staff are unable to give formative feedback on student learning to the extent they may wish".

Hence, the solution might be in seeking to change the mode of assessment and making use of IT as "there are potentially great time savings to be made by the automatic marking of students' work" (Bull & McKenna, 2004, p 7). Furthermore, it has been made clear that applying automated marking could enable academics to provide their students with timely, accurate and individual feedback as "assessments which are marked automatically can offer immediate and evaluative statistical analysis allowing academics to quickly assess whether their students have understood the material being taught, both at an individual and group level" (Bull &McKenna, 2004, p.6).

2.2.7. Assessment Acceptability

Computer Assisted Assessment or CAA (also known as: Computer Aided Assessment, Computerized Assessment, Web Based Assessment, Computer Based Testing and Computer Based Assessment) refers to using the computer to deliver, mark and analyze students' examinations (Bull & McKenna, 2004; Seale, 2002).

Today, the use of the computer for assessment is an innovation which is causing much controversy over the appropriateness, efficiency and applicability of IT in assessing students. Moreover, this innovative assessment mode is facing many challenges because assessment designers still find it difficult to further develop and implement computerized assessment (Drasgow & Olson-Buchanan, 1999).

There is sluggishness in CAA development and successful implementation as argued by Bull (1999a) which is mainly due to several cultural and organisational barriers such as higher education institution's lack of commitment to implementing CAA. Bull (1999a) notes that "rarely have departments, faculties or whole institutions shown a commitment to implementing CAA"

(p. 123).

Although a new technique or method of assessment could be defined as innovative assessment, Harris & Bell (1990) point out that "it is not the actual methods or tools of assessing which we believe should be changed in many cases, rather the underlying philosophy and the aims of their use and application" (p.97).

In fact, there is today a flourishing literature about CAA's capacity to improve assessment, thus improving students' learning (Bull, 1999a; Bull & McKenna, 2004; Chalmers & McAusland, 2002; Drasgow & Olson-Buchanan, 1999; Ricketts & Wilks, 2002a, Seale, 2002). Bull (1999a), for example, notes that "CAA can provide academic staff with the opportunity to review and refine their assessment strategies holistically"(p.124).

However, as it is the case with any innovation, there are many sceptics who resist changing the way students are assessed. This may be due to various emotional or cultural reasons like complacency and lack of motivation which justifies satisfaction with the existing assessment mode. Another reason for rejecting the adoption and implementation of CAA could be the lack of knowledge about its benefits, which results in developing negative attitudes towards it. We should also note that applying any innovative assessment method implies changing assessment culture and as a result this process is always accompanied by much resistance. Therefore, a great deal of effort from all parties involved in the process (like stakeholders, academics, researchers) is needed if CAA is to be applied and embraced, for "changing assessment procedure is often more difficult than the process of assessment itself" (Brown, Bull & Pendlebury, 1997, p 222).

One of the greatest advantages of CAA is its potential to make use of various assessment methods to enable the assessor to incorporate graphical and visualisation effects like diagrams, maps, animated images, sound, or video clips in the construction of an objective test. In addition, CAA can be useful for both assessors and learners alike, as it ensures that reliable and objective marking is taking place especially with large student populations, as well as providing instant analysis and assessment of results (Bull & McKenna, 2004).

However, the two biggest challenges facing academics when applying computerised objective testing are: First, the objective test is presumed unable to go beyond the assessment of knowledge by addressing the application of this knowledge in a critically thinking way. Second, sparing the time needed for constructing effective computerised questions to assess these higher order thinking skills. In fact, the biggest advantage of CAA (the possibility to incorporate multimedia) does create a major difficulty for test designer or assessors, for it is time-consuming. That is, academics will face several technical challenges as incorporating graphical and visualisation effects does require much time and effort which will be added to the initial effort and time spent on constructing effective questions that could be used to assess higher order thinking skills (Front Load). For these reasons, adopting CAA may not be always appreciated by academics. Hence, the need to explain the potential of computerised testing and analyse its benefits (like: generating timely feedback, reliability and creating items' banks) over traditional testing modes.

Academics will also need to undertake adequate training to get a thorough grasp of CAA's potential before being urged to apply it. This view is stressed by McKenna & Bull (2000) who point out that "the swiftness of technical and pedagogical developments in CAA necessitates that both novices and experienced practitioners acquire and maintain new skills. A strong programme of staff development should help ensure competence in the following areas: use of CAA software, training in construction of appropriate questions, invigilation of CAA exams, test design, embedding CAA within an existing module, selecting a mixed range of assessment methods and evaluating CAA" (p. 28).

2.2.8. Assessment Cost

Another obstacle that might hinder the application of CAA is the cost factor It is obvious that applying computerised testing is costly therefore resources and funding must be provided to ensure the successful application of CAA This involves the provision of extra funds for teaching, training, providing venues, providing assessors and assessment designer, getting adequate hardware and software Hence, it is only by successfully managing the whole assessment process and securing public acceptance and support that applying CAA is made feasible and successful.

Ricketts et al., (2003) also provide an illustration of the items related to the cost of implementing CAA as well as the potential benefits of such a process for the university, the staff and the students They point out that even though the implementation of CAA would raise money worries as to the provision of adequate software and robust hardware besides the money needed for the support staff, the maintenance of the server, the development of the staff and the additional load on IT infrastructure, still universities will gain substantial benefits in return. This includes the compensation for the lack of staff recourses (time), substantial reductions in the cost of paper and printing, a better retention of students and an increased cross-departmental working In addition, while all that the staff will have to do is spare "time to learn the system", they will in return draw numerous benefits such as " no marking; easy recording of marks; better information on students[and on] assessment; an additional mode of assessment and improved staff-student contact" (Ricketts *et al.*, 2003, p.333).

Finally, as Van Der Vleuten (1996, 62) points out that "Extending assessment technology towards maximal fidelity and its planned educational use will be the challenge for the future".

2.3. Constructive Alignment

One of the advantages of CAA is to enhance learning by enabling academics to achieve 'constructive alignment'. As most of higher education institutions use computers in teaching and learning, so using computers in assessment will align the educational process and will enhance learning. Constructive alignment is a theory of learning which builds on the premise that the learner constructs his/her own learning according to the teaching activities adopted by their academics (Biggs, 1999; Biggs, 2003). This theory also stresses the need to urge students to assume responsibility for their own learning by declaring the Intended Learning Outcomes (ILOs) of a course or a module, and aligning them with the teaching methods and assessment procedures so that students construct their learning by structuring their learning activities in a way which leads to meeting assessment requirements or criteria.

This means that academics should define the goals or the Intended Learning Outcomes (the specifications of what students should be able to do following a course) and then create a supportive learning environment by selecting appropriate teaching activities which are likely to enable students to attain the pre-set ILOs, and aligning both with assessment procedures.

The consistency of the system is meant to enhance learning as noted by Dochy *et al.*, (2007, p.87) who argue that "this alignment might significantly increase the power of assessment as a stimulus and a tool for learning". Hence, CAA has to be aligned with both the ILOs of the course or the module and the teaching practices carried out by academics, for it is otherwise unfair to teach students in one system and assess them in another (Brown, Bull & Pendlebury, 1997). Biggs (1999) also highlights the strong correlation between teaching and assessment by stressing the need to get the Intended Learning Outcomes of a course or a module, the teaching practices and the assessment criteria aligned and consistent if we wish teaching to be constructive, for the consistency of the system is meant to enhance learning The same view is expressed by Dochy et al., (2007, p 87) who point out that " this alignment might significantly increase the power of assessment as a stimulus and a tool for learning". Moreover, in the Applied Colleges in Oman, there has been a considerable change in the learning and teaching activities and a substantial emphasis on the practical aspects of learning rather than solely focussing on the theoretical ones. Therefore, the most important thing is to get the assessment methods aligned with the learning objectives, because if we change the way of learning without changing the assessment methods, learning will never be enhanced or improved (Elton & Johnston, 2002). So, as Omani higher education institutions implicated e-learning this implies that e-assessment should also be applied to align the assessment method with the teaching process.

This implies that academics need not only to communicate the ILOs of a course or a module to their students but also adopt teaching methods and practices which are aligned with the assessment methods to optimise their students' learning. Constructive alignment, then, advocates transparency in declaring ILOs and consistency in aligning them with both teaching practices and assessment methods.

2.4. Advantages and Disadvantaged of Computer Assisted Assessment

Since the most important aspect of assessment is quality and not quantity, we might assume that CAA does not necessarily mean more assessment administration. We might also assume that central to applying CAA is the tendency to reduce the work overload faced by academics while still preserving the primary aim of assessment which consists of enhancing learning.

It is obvious that CAA can cater for this by staging assessment more regularly throughout the academic year, which in turn, prompts more regular study behaviour. It has also been argued that CAA can improve students' learning in various other ways thanks to the considerable advantages it has over traditional assessment practices (Bull & McKenna, 2004). These advantages are mostly noticeable today when academics have to cope with a rapidly-increasing number of students in higher education. Bull (1999b, p.3) notes that "well-designed CAA can be used to enhance the student learning experience, expand assessment processes and provide efficiency gains for academics and support staff to keep pace with rising student numbers".

The advantages of applying CAA include (Bull & McKenna, 2004):

a. Motivating students to learn and encouraging them to practise skills.
b. Increasing the frequency of assessment, objectivity and consistency.
c. Increasing feedback for teachers and students.
d. Varying assessment methods.
e. Decreasing marking loads and aiding administrative efficiency.

We should note that although these advantages are presented as discrete points, they are, nonetheless, overlapping, inter-related and serving the same aim which is enhancing students' learning. The various advantages presented by CAA can be perceived as revolving around three major inherently related objectives: saving time and effort, boosting motivation and enhancing learner autonomy. The three objectives will be analyzed respectively in a way that reveals their interrelatedness and their capacity to foster and enhance learning.

The benefits of immediate scoring and generating feedback make computerized assessment easy to administer and easy to assess without adding to the work overload which is already faced by busy academics. This both encourages academics and enables them to increase the frequency of carrying out formative assessment since both data collection and analysis are carried out by the computer. Therefore, lecturers will only have to utilize the available results in a way which should lead to improving students' learning.

Besides, timely feedback can also be exploited by students who may find it developmental and formative since it is generated while there is still ample time to reflect and act upon their performances in order to improve them.

Another point is that assessment is the major driving force behind learning, for students might not be motivated enough to engage actively in learning unless they realise that what they have to learn will be subject to assessment (Brown, 2004, Gibbs & Simpson, 2003). Brown (2004, p.81) points out that "students can and do ignore our teaching; however, if they want a qualification, they have to participate in the assessment processes we design and implement". Also Chalmers & McAusland (2002, p.2) note that "if topics do not have some sort of assessment procedure, students may not apply themselves to the topics with a preferred vigour".

Also CAA can provide a substantial incentive for students to learn, especially as objective testing - being the most commonly used form of CAA (CAA Centre, 2002; Crisp & Ward, 2008) - offers huge capacity to cover a large portion of a course or a module. Thus, as pointed out by Brown, Bull & Pendlebury (1997, p.7), the assessment style can have a big impact on students' learning; hence, the extrinsically motivating role of CAA.

Another advantage of CAA consists in its tendency to foster learner autonomy by enhancing self-assessment practices which tend to "enable students to gauge their own understanding of the material and learn from their mistakes via the available feedback" (Stephens & Curtis, 2001, p.7).

This means that students' role will change as their commitment to and engagement in learning increases. This will, in turn, result in producing dedicated and active learners who assume responsibility for their own learning, for self-assessment implies "the involvement of students in identifying standards and/or criteria to apply to their work and making judgments about the extent to which they have met these criteria and standards" (Boud, 1995, p.12).

Once again, we come to realise the tremendous role of CAA and its substantial capacity to produce students who are not only motivated but also committed, independent and deep learners. Although much focus has been laid, so far, on the significance of CAA in enhancing regular and sustainable learning behaviour, for it greatly lends itself to formative assessment practices , it is nonetheless important to note that CAA can also be deployed for diagnostic or summative purposes (Bull et al., 2002).

CAA presents many benefits to the academic staff by enabling them to receive feedback during or immediately after any formative assessment process. This feedback serves to help evaluate the effectiveness of the course content and design. It also enables them to monitor students' progress more easily through the frequent use of assessment, enabling assessors to cover a wider range of topic areas and enabling tutors to devise and assign tailored remedial activities to students in the light of the feedback they have got, which optimises students' development and fosters their learning.

Chalmers & McAusland (2002) and Seale (2002) also present several administrative and organizational benefits which include: saving time spent on supervision, invigilation and marking of assessment, reducing the rate of subjectivity and human error, for "computerised marking is not prone to human error" (Seale, 2002, p. 3), saving resources particularly when assessing expanded student numbers; saving time by automatically generating statistical evaluation; assuring the successful integration into, and coordination with students' records and the university information and management systems, and substantially reducing printing costs, especially when tests are updated or altered

Many other researchers advocate the numerous pedagogical and organisational benefits of CAA by pointing to the advantages it has over traditional assessment modes particularly in boosting students' motivation. For example, many researchers argue that when attempting to implement CAA in any educational institute, is necessary to evaluate the benefits that institute will gain in comparison to other methods of assessment and to clarify the CAA possibility to enhance the educational process in relation to feedback and motivation (Bull, 2002). The same view about CAA's tendency to boost students' motivation is noted by Ricketts & Wilks (2001, p.418) who state " A change in assessment could increase motivation".

Most studies also stress the importance of timely formative feedback. Bull (1999a) stresses that central to CAA is its potential to provide focused and timely feedback which she considers very developmental for both parties, academics and students in addition to its being a substantially motivating factor which drives learning. She writes "perhaps the most valuable benefit of CAA is the ability to provide focused and timely feedback to students and staff. Feedback can be used to direct future learning, motivate students to investigate other resources and identify students who need additional support" (Bull, 1999a, p.123).

The same view is also stressed by Brown, Race & Bull (1999) who advocate CAA's capacity to provide students with detailed formative feedback on their learning in a more efficient way than is usually possible with traditional assessment modes. Similarly, Ricketts & Wilks (2001) argue that if feedback is not detailed enough, it would have little formative value. Bull (1999a, p.125) also stresses the speed of feedback as a determinant of improvement. She writes " the quality and speed of feedback which students' receive can be enhanced by CAA and the extent to which academics are aware of their students' progress and deficiencies may be increased". Other studies highlight the substantial potential of CAA to enhance learning by stressing its role in facilitating and increasing the frequency of formative assessment administration.

However, while Lawson (1999) shares the same view and speaks about CAA's role in providing an " opportunity for repeated practice", Ricketts & Wilks (2002a) add that CAA's benefits are meant to culminate in improving summative assessment. Ricketts & Wilks (2002b, p.475) argue that if CAA is used for formative assessment, students' performance in summative assessment improves. They state "One of the benefits claimed for computer-based assessment is that it can improve student performance in summative assessments". We should note, however, that using CAA for formative assessment purpose implies integrating it in the teaching/learning process. So it may help Omani students to get a benefits of CAA in develop their study skills Charman & Elmes (1998) and Sly & Rennie (1999) even provided evidence that students perform better if computer-based assessment is integrated in the learning process. Although feedback is central to learning, what really matters is the quality of feedback. Hence, we might assume that the feedback provided by computer-based assessment, however important it is, is still subject to some limitation. That is, CAA has a tremendous capacity to provide students with non-judgemental or reliable timely feedback but nonetheless may fail to give them adequately discriminatory or differentiated type of feedback. Nevertheless, it is argued that with more effort, it would be possible to generate intelligent feedback, but this would be time consuming and expensive However, we should never underestimate the benefits of CAA altogether, for it has a substantial capacity to improve students' learning, especially if it is used for formative assessment purposes (Charman & Elmes, 1998; Mackenzie et al., 2004; Sly & Rennie, 1999).

Academic staff can also benefit tremendously from getting instant feedback about their students' responses, which enables them to reconsider the curriculum and the teaching methods, review and refine their assessment strategies and spot students who need additional support (Bull, 1999a; Bull & McKenna, 2004). It might also be assumed that computerised assessment saves academics' time spent on marking scripts and, thus, enables them to spend time on improving learning and teaching methods. That is, instead of spending that time on marking after the administration of a test, academics devote this valuable time to both devising efficient assessment before the administration of a test and preparing remedial plans based on the feedback they get after the assessment has taken place. Brown, Rust & Gibbs (1994) cited in Brown (2004, p.83) say " we cannot simply expect our students or ourselves to just keep working harder and harder; where possible we must make best use of the available technologies to make assessment more efficient"

In spite of the great interest in technology by the new Omani generation, and students' strong desire to embrace the computer world, no study has so far investigated the potential effects of applying computerised assessment on Omani higher education institutions or investigated students' potential attitudes towards this novel assessment mode. So, the transition to CAA in Oman would be relatively unpredictable unless it is carefully planned. However, as the use of computers in teaching and learning have been emerging in many of Omani institutions, so it would be easier and safer to promote implementing computerised assessment as it assumed that students become familiar with the use of the computer and, thus, they could develop positive attitudes towards computer-based assessment as recommended by Barnett (1995) .

CAA is not only considered as a practical solution to deal with expanding students' numbers in higher education institutions but also an innovative assessment method which could offer numerous pedagogical benefits which tend to optimise students' learning by improving teaching and assessment methods. So I like to use the diagram below (Figure 2) which summarizes and illustrates the various pedagogical benefits of CAA and how it serves to enhance students' learning.

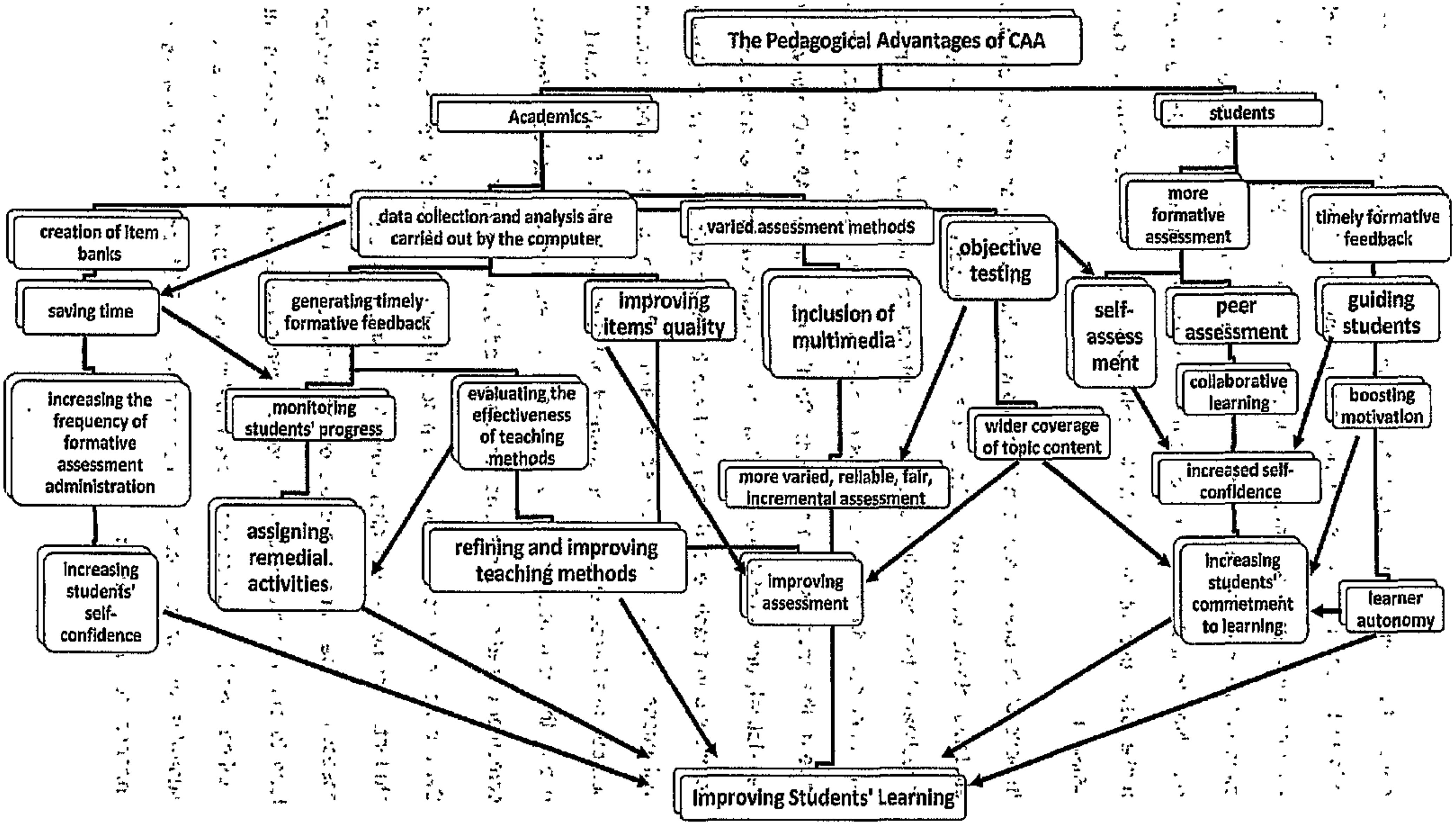


Figure 2. Pedagogical Advantages of CAA

As for the pedagogical disadvantages of CAA, they mostly relate to two main issues: quality assurance, and the initial set up time and learning curve. Although these two issues are seemingly distinct, they are nonetheless closely interrelated. With regard to quality assurance, the objective test can neither assess higher order thinking skills nor students' communication skills. This, in turn, requires that academics in Oman and support staff be well trained in term of IT and assessments design so that they could meet the quality assurance requirements of devising appropriate and efficient adequate objective tests. Another point is that invigilators should also be trained in IT and assessment design, for students may raise different issues in computerised assessment from the ones raised in the paper-based assessment modes.

The challenging CAA disadvantages that might face Omani institutions relate to the need to ensure there is a high level of inter- and intra-departmental cooperation and coordination between academics, support staff, administrators, technologists, etc. Otherwise it would be extremely difficult to ensure the successful implementation and the smooth functioning of CAA, particularly under exam conditions (Chalmers & McAusland, 2002). The need is to ensure the availability of lab space and robust computers, so that large numbers of students, possibly belonging to different departments, may be assessed simultaneously (Chalmers & McAusland, 2002; Seale, 1999).

As for the 'disadvantage' which is related to the initial setup of CAA, it is argued that since assessment is integral to teaching and is considered as a part of students' commitment to learning (Brown, 2004; Brown, Race & Bull, 1999; Gibbs & Simpson, 2003), changing assessment methods should be pedagogically rather than resources driven. That is," the extent to which [CAA] will enhance student learning should be the paramount criterion, rather than how much time and other resources it will save" (Chalmers & McAusland, 2002, p.3). Accordingly, even if CAA might be costly to set up, its substantial capacity to offer numerous pedagogical benefits (such as: accounting for data collection and analysis; providing timely and formative individual feedback; evaluating the effectiveness of teaching methods and practices; encouraging student self-and peer assessment; boosting students' motivation; varying and improving the authenticity of assessment and creating items banks) should not be overlooked.

As for the time saving variable, it should be stressed that although the initial setup of CAA might be time-consuming, substantial time saving could be achieved in the long run. Bull & McKenna (2004, p.7) note that "there are potentially great time saving to be made through the automatic marking of students' work". Moreover, saving academics' time would help to meet a fundamental pedagogical requirement as this time could be invested in improving students' learning by increasing the frequency of formative assessment administration. Saving academics' time would also be invested in monitoring students' progress, for "assessments which are marked automatically can offer immediate and evaluative statistical analysis allowing academics to quickly assess whether their students have understood the material being taught, both at an individual and group level" (Bull & McKenna, 2004, p.6). Another point is that saving academics' time would help to improve assessment practices (as illustrated by the diagram figure number 2 on page 52)

So, the demand for saving academics' time should not be understood as an attempt to rid academics of part of their duties but should, instead, be perceived as an attempt to spare the time which they should be investing in improving students' learning. Hence, what is actually sought is time redistribution and not time saving as such. This view is pointed out by Bull & McKenna (2004) who argued that "using CAA can demand a cultural shift in terms of time invested in assessing students' learning. Academics need to invest time prior to the event, rather than after it" (p.7).

Thanks to the prevalence of computer literacy and 'e-culture' today even in Oman, acquiring the adequate and sufficient IT skills required by CAA will not make a real challenge for academics, invigilators or students. We might even stress that acquiring adequate IT skills should be further encouraged so that students gain the essential competencies which they would need after graduating from university. This view is asserted by Bull & McKenna (2004) who point out that "in nearly all areas, CAA may improve the authenticity of assessment, since work with computers will almost certainly play a part in what students do during and after leaving university "(p.141).

Since the computer has become widely used in teaching (CAL) in Omani institutions, why should it not then be used for assessment purposes, especially if that is going to ensure the consistency of the system? Bull & McKenna (2004) assert this view by pointing out that "if computers have a role in teaching and learning, it seems appropriate that they should also be part of assessment practices" (p.11).

In real life, the society, and particularly those involved in or affected by any kind of change, are influenced by a vast range of values and beliefs, which may lead them to develop sceptical views or even a tendency for rejecting any innovation (LTSN, 2002) .This means that changing assessment methods (in Oman) may be faced by scepticism or rejection, for "in assessment, one has to deal with opinions, sentiments and traditions of teachers, students and institutions." (Van

Der Vleuten, 1996, p 54). In order to get people to embrace CAA, academics, students and stakeholders should be provided with sufficient information about this new assessment method as well as its benefits. Van Der Vleuten notes that " the extent to which an assessment procedure is accepted by the people *involved in the assessment is a crucial element for consideration"* (1996, p.54). In the Omani context, students have diagnostic, continuous formative and summative assessment during the whole year. Also students in the foundation year (first year) have to sit for diagnostic English assessment prior to their enrolment in the universities so that they are categorised according to their proficiency level. Students also sit for formative assessment tests which are meant to help to improve their English level during the foundation year before taking summative assessment according to which their English language proficiency and mastery of English is assessed to determine whether or not they can proceed on to university Hence, it would be useful to consider the pedagogical and administrative/organisational advantages and disadvantages of CAA and what it can offer to the Omani higher education institutions.

Society's role should not be ignored, either, as it plays a significant role in the acceptability (or perhaps rejection) of CAA. Dochy *et al.*, (2007) argue that that indications show that students do not develop positive attitudes towards a new *assessment procedure or opt for it unless they are given sufficient information* about it and its benefits beforehand. So, decision-takers will have to advertise and publicise CAA and give value to its benefits if they want to ensure that it is accepted and embraced.

Finally as asserted by Sternberg (2007), assessment can only be understood within the cultural context in which it is carried out. Therefore, the next chapter will address the assessment in Oman within Omani culture to get a clear picture of assessment in the Omani context.

Chapter Three: Culture, Education and Assessment in Oman

3.1. Introduction

The Sultanate of Oman is an increasingly and rapidly developing country, and both the Omani culture and the social structure have been affected by the economic and social renaissance that started in 1970. To cope with these changes, the educational system had to be changed and improved

It is evident that student assessment in higher education cannot be thoroughly understood outside the social and cultural contexts in which it is carried out This view is asserted by Thorpe, Edwards & Hanson (1993, p.2 Cited in Al-Alawi, 2004, p.13) who argue that "teaching and learning activities, including assessment practices are expressions of culture whatever form they take". And, Oman is no exception, for the Omani educational system is influenced by the tacit Omani culture

The purpose of this chapter is to describe Omani culture and the educational system adopted and applied in Oman in order to get a clear understanding of the factors which influence education, and particularly assessment in higher education

3.2. Omani Culture

Brosnan, Scheeres & Slade (1999) "highlight six different approaches in relation to culture. These approaches are: ethno specific, EEOC/anti-discrimination [Equal Employment Opportunity Commission], socio-historical, linguistic, psychological/interpersonal and productive diversity". It is of great importance to take the effects of the cultural and socio-historical factors into consideration in the context of this research project to get a good understanding of the local situation and how it differs from western norms.

As in other nations, Shaw (1996) argues that the Arab culture involves many variables such as gender, religion, regional differences, attitudes, politics, and that these variables are closely related to the history of the region. Moreover, the Arab culture has its own perceptions of the nature of knowledge, learning and teaching approaches which are different from western views and perceptions (Shaw, 1996).

It is should also be noted that, as argued by Al-Alawi (2004), the word 'culture' has different definitions; however, we could assume that the one provided by National Languages and Literacy Institute of Australia (NLLIA, 1995) could be the most suitable definition in the context of our research. According to NLLIA (1995, p.14) "there are no fixed boundaries to culture and cultures are always changing. Any individual lives in and between many different cultures: the culture of the work place; the culture of educational institutions; culture as ethnic background; culture as aspiration, interest or inclination. In this sense, all our cultures have multiple layers, each layer is in a complex and dynamic relation to the others" (cited in Al-Alawi, 2004, p.14).

Geographically, the sultanate of Oman covers 309,500 square kilometres, which makes it the third largest country in the Arabian Peninsula (*Oman Yearly Book*, 2009/10, p.12), whereas UK is approximately 245,000 square kilometres (www.enotes.com/topic/United_Kingdom). The estimated resident population of the UK was 61,792,000 in mid-2009 (*www.statistics.gov.uk*), while the total population of the Sultanate of Oman according to the 2003 census was 2,331,391 (*www.omancensus.net*).

Being 'huge', Oman is not a densely populated country (about 8 persons per square kilometre) the Sultanate also has a coastline which stretches for more than 3,165 kilometres (Ministry of Tourism, www.mot.gov.om). It shares borders with the Republic of Yemen to the southwest, the Kingdom of Saudi Arabia to the west and the United Arab Emirates to the north (*Oman Yearly Book*, 2009/10, p.12) Oman also has special demographic features as only 14% of the Omani population live in cities and towns (Al-Alawi, 2004).

The Sultanate of Oman is administratively divided into nine Regions and Governorates (Figure 3): Muscat, Dhofar, Musandam, Al Buraimi, Al Batinah, Al Sharqiyah, Al Dhahira, and Al Dakhiliah. Each Governorate / Region is formed of "Wilayats" (totalling 61) (Ministry of Tourism, www.mot.gov.om).

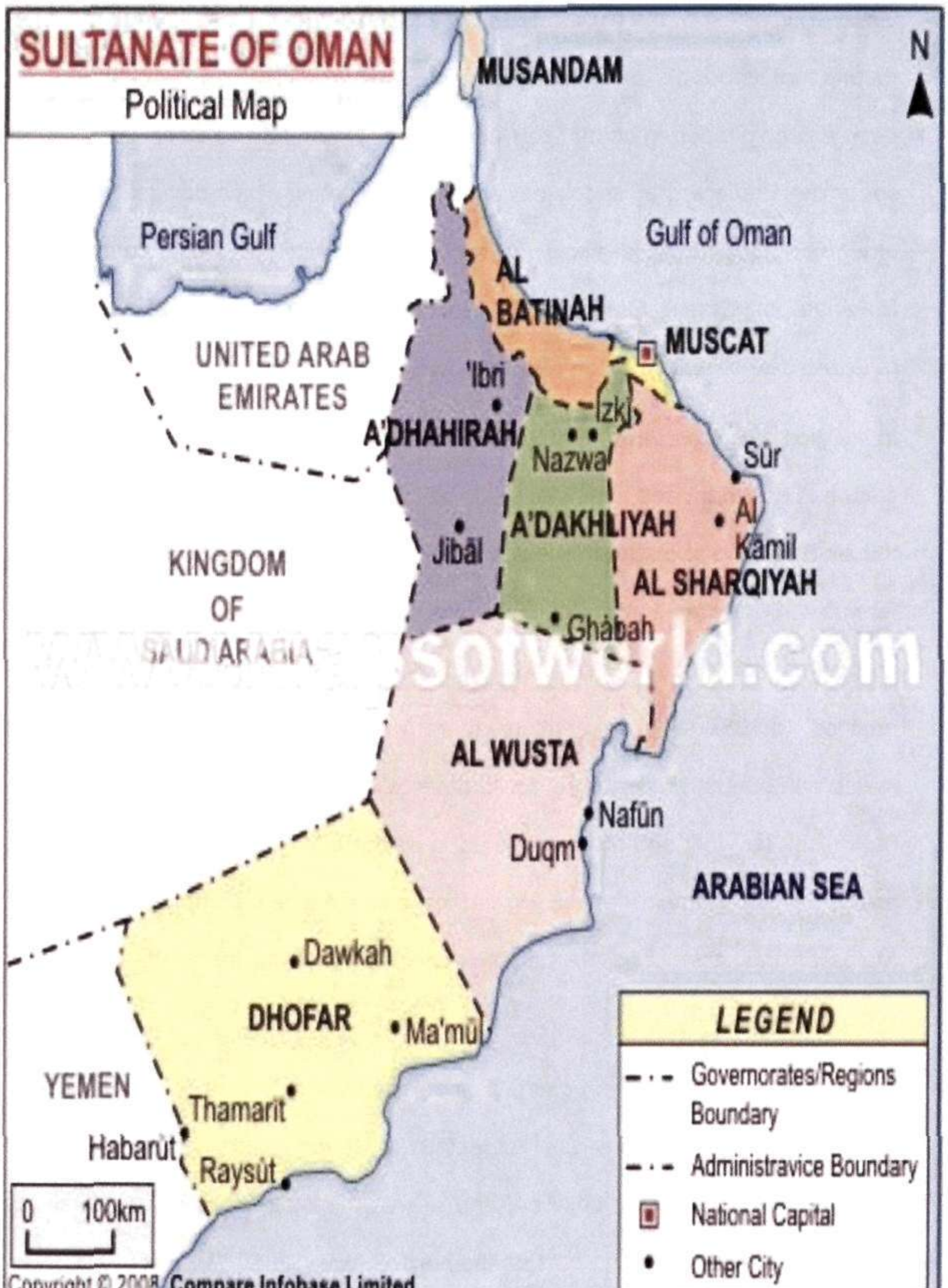


Figure 3. Oman Map

Adopter from www.mapsofworld.com

Geographically, too, the Sultanate of Oman is situated in the Gulf region, which means that its people share many beliefs, habits, traditions and values with the people in neighbouring countries. Nonetheless, Oman has its distinctive cultural specificities reflected by its folklore and national dress, which distinguish it from other Gulf Countries (Al-Alawi, 2004). Even within the same country (the Sultanate of Oman) Omanis have various regional specificities which are reflected by different types of folklore as well as habits and traditions.

In addition, like other Arab countries, Oman is a Muslim country. Hence, Islamic culture is dominant and affects the different facets of intellectual and social lives there. Speaking about how Islamic culture is the dominant culture in Oman, Al-Rawas (2001) states that "it has been suggested that the Islamic religion dominates social behaviour and functions as the source of humanitarian motives" (p.198). In Islamic culture, there are specific roles which are set for males and others which are set for females. While males are encouraged to go out to look for work and are considered to be the family bread winners, females are usually discouraged from going outside and are given a different social role. Males on the other hand are usually expected to inherit their father's craft or profession.

As for females (woman and girls), their role is usually limited to taking care of the home, and mothers are expected to teach their daughters house management and raising children, and are as a result expected to remain at home and preferably not go outside the house.

However, nowadays males and females are equal in terms of rights and duties. Also, they have the freedom to choose the way and the style they wish to live without breaking the social and Islamic way of living.

3.3. Education in Oman

3.3.1. Development in Education Sector

Since the accession of Sultan Qaboos bin Said to power in Oman in July 1970, a number of considerable changes have taken place in the educational system in Oman. Perhaps the most remarkable improvement has been in the number of schools which has substantially developed. Before 1970, there were only 3 modern boys' schools in the entire country with a total number of 909 students and 30 teachers. At that time the Omani educational system consisted mainly of the 'Quran schools' which were wide spread throughout the country and run by the ' Imams' (men who lead the five prayers in a mosque) or 'sheikhs' (men who have good knowledge of the teaching of Islam) these 'schools' were called 'kuttab' or 'madrasa' (Al-Alawi, 2004).

In the past, attending government schools was restricted only to males while girls were denied the right to attend schools. For girls however, up to a certain age, they were allowed to attend 'Quran schools so that they could learn the teachings of Islam, and above all to learn about how to read the holy Quran in a proper way. Then, girls had to remain at home and learn more about house management.

It is difficult to imagine that in 1970, a country with a population of approximately one million inhabitants had only three schools and nine hundred and nine students, which means that it was almost universally illiterate. However, Oman has incredibly changed and Omani people have surpassed the hard period preceding 1970 and moved to an enlightenment period under the guidance and leadership of his Majesty, Sultan Qaboos bin Said.

Thanks to the fast development of the educational sector after 1970, the number of government schools has risen tremendously, reaching 1050 schools in the 2008-2009 school year. This substantial increase in the number of schools has been paralleled by a similar increase in the number of students (which reached 541,436), and the number of teachers which has grown to 43,672 in the same school year. Besides government schools, there were 339 private schools in the 2007/2008 academic year, attended by 55,000 students and 4000 teachers according to the statistics provided by the Ministry of Education (MOE) in 2008/2009 and cited in Oman yearly book 2009/2010 (www omanet.om). The Ministry of Education also supervises a total of 114 Holy Qur'an schools, as well as 33 international schools for the children of expatriates according to the Oman yearly book (2009/2010). These schools, both government and private, provide education for both male and female students at the primary, preparatory and secondary levels. In all Omani government schools, Arabic is the medium of instruction while English is taught as second language. However, in higher education institutions, English is considered as the first language, and also the medium of instruction in most Omani colleges and universities

In fact, since, the outset of the renaissance schools have quickly started to spread all over the country, so the Omani community started to assimilate the benefits of the radical changes in a fast way, and gradually accommodate themselves to the changes of the new era. For the first time ever, Omani girls were allowed to attend government schools and were gradually admitted into higher education institutions. However, their admission into college and universities was initially restricted to certain streams. The tendency was principally to get girls to join educational rather than scientific streams.

Nevertheless, starting from the early nineteen-nineties, female students have started to enrol for scientific streams along with male students although the female enrolment rate in science is still lower than the male's as the majority of females prefer to join the educational sector after graduation. Girls are now treated equally with boys in terms of education. Female students can even study abroad today on an equal footing with male students, so it has been possible for them to join international universities and have the complete freedom to choose their desired colleges and subject areas.

As a result, girls can now apply for and take various jobs according to their qualifications. As evidence of this, I have myself obtained my first and second degrees from universities situated outside Oman.

So, ever since his Majesty Sultan Qaboos bin Said started ruling the Sultanate of Oman, the educational sector has started to improve in a remarkably fast way; no wonder as this particular sector received his Majesty's complete attention and was perceived as a top priority sector in order to build a developed modern country.

It is important to note that since 1970 education in Oman has passed through two different phases. The first phase lasted from 1970 to 2000 and the second phase from 2000 till today. Thus, there are currently two different educational systems which are applied in Oman. In the old system, there are three educational stages; primary, preparatory and secondary schooling. Primary school education usually starts when the child is six years old. Having passed the sixth primary grade examination (usually between the ages of twelve and fourteen), the pupil moves on to the preparatory school. After spending three years at the preparatory school and passing the general preparatory certificate examination, the pupil is then admitted to the first grade of secondary education (between the ages of fifteen and seventeen) which lasts for three more years. This system is called General Education whereby the curricula and course books are focused mainly on the Omani context; that is Omani history, geography, society and culture. In the general education system, grade eleven students (those age between16 and 17) have to choose either a science or an arts stream. On the completion their secondary education, successful students can move on to the advanced training of specialized colleges or enter Sultan Qaboos University (SQU).

At the beginning of the 1998-99 academic year, the Ministry of Education (MOE) start a new educational system which is called Basic Education which consists of two cycles followed by a two-year secondary school period. The first cycle comprises grades one to four where children start going to school at the age of six. As for cycle two, it lasts for six years (grades five to ten) then, students reach the secondary school where they spend two years (grades eleven and twelve). This means that the basic education phase covers ten years and the secondary education phase covers a period of two years. As planned, this new educational system (Basic Education) is gradually replacing the old one. Under the new system, there is a greater emphasis on other subjects such as science, maths and computing. English is also of much greater importance and is introduced from the first year of basic education that is, from grade one, unlike in the old system in which it starts from the fourth grade.

According to the Ministry of Education, the new system has been introduced gradually starting with seventeen schools from different regions of the sultanate. After one year; that is during the 1999-2000 academic year, the basic education system was further introduced to twenty-five more schools, it was also introduced in some private schools. In the 2000-2001 school year, the number of schools implementing basic education rose to 101, and according to the educational statistics year book, there were 507 schools implementing this system in 2005-06 while 539 other schools were still under the old system (Ministry of Education, 2006). To meet the requirements of the new educational system, new learning resources centre have been set up in all government schools. These resource centres consist of equipment (including computers) as well as audio-visual learning and teaching aids. According to the Ministry of Education, the number of the schools implementing basic education will gradually rise in the following years till the new system is implemented in all schools (Ministry of Education, 2006).

Changing the educational system in Oman has been accompanied by changing the curricula, so the new syllabuses and text-books are now open to the outside world and not solely focused on the domestic context as they used to be.

The assessment system has also substantially improved, for new perceptions and approaches to assessment have been adopted and implemented. As for the nature of assessment which is used in the General Education system, it is rather summative, for focus is mostly on the tests that are conducted at the end of each semester. Now, under Basic Education, focus is laid on continuous assessment which is carried out throughout the academic year for formative rather than solely summative purposes. Under the new assessment system, students' performances are recorded on pre-set assessment sheets and monitored throughout the whole school year so that remediation measures are taken while there is still time for the students' improvement.

Unlike the UK where coeducation is common, there is still a separation between boys and girls in schools. That is, there are schools which are attended by girls and ones which are attended by boys, particularly in the General Education system. However, the situation is slightly different in the Basic Education system where the separation of boys and girls starts after grade four. In other words, coeducation is applied in cycle one while the separation of both genders starts in cycle two. In cycle one, both boys and girls are taught by only female teachers however, starting from cycle two, boys and girls are taught by male and female teachers respectively

3.3.2. Development in Higher Education Sector

In 1986, the first university (Sultan Qaboos University) was set up and opened to the public in the Sultanate. Today, this university (SQU) offers advanced courses in various disciplines and also provides the needed research facilities for researchers and scholars as well as post-graduate students who are carrying out studies both at Master's or Doctoral levels. Some private universities offer post- graduate degree programs at the Master's level only. The number of Omani higher education institutions on 2007/2008 was 57 (*Oman Yearly Book*, 2009/10). The total number of students' studying inside and outside the country reached 78,930 according to Oman yearly book 2009/2010.

Higher education institutions in Oman, including SQU, are funded by the Omani government The academic staff members who teach in these universities and colleges consist of Omani and non–Omani (expatriate) academics; nevertheless, until now the majority of those academics are still non – Omani.

However, the situation in higher education is different as coeducation is generally prevalent (mixed classes) and students are taught by either male or female academics, except for some disciplines or some private colleges where male and female students are taught separately. As for the Omani higher education institutions, they consist of: Sultan Qaboos University, The Colleges of Applied Science, private universities and colleges, Health Institutions, Technical and Industrial colleges, Institute of Sharia Science, and the Institute of Banking and Financial studies (www.mohe.gov.om). The system of higher education is governed by a number of different ministries and government bodies in Oman (Al-Lamki, 2002; Al-Lamki, 2006). The table number 2 below shows the number of government institutions and private ones according to the Ministry of Higher Education, Ministry of Health and Ministry of Manpower websites.

Ministries and Government Bodies	Institution	Number
Council of Higher Education	All institutions of higher education	
University Council	Sultan Qaboos University (SQU)	1
Ministry of Higher Education (MOHE)	Applied Science Colleges	6
Supervised by MOHE	Private Colleges	19
Supervised by MOHE	Private Universities	6
Ministry of Manpower	Technical Industrial Colleges	6
Ministry of Health	Health Institutes	15
Central Bank of Oman	Institute of Banking and Financial	1
Ministry of Awqaf and religious Affairs	Institute of Sharia Science	1
Ministry of Tourism	Omani Academy of Tourism and Hospitality	1
Royal Office	Royal Guards College	1

Table 2. Higher Education Institutions Governance in the Sultanate of Oman

The Ministry of Higher Education (MOHE) was established in January 1994 and oversees six education colleges. In addition, MOHE is also responsible for all private colleges and universities.

During the 2005-06 academic year, the ministry converted five of the six colleges of education into Colleges of Applied Sciences. After two academic years the sixth college was converted to an Applied College as well These six Applied Science Colleges, which are situated in six different regions (Nizwa, Ibri, Sur, Sohar, Rustaq and Salalah), are all overseen by the MOHE

These colleges offer 5-year academic programs, including a one-year foundation course. The foundation year is followed by a four-year Bachelor's degree course in Information Technology (IT), International Business Administration, Design, or Communication studies.

As the medium of instruction is English, the one-year foundation course is devoted to consolidating students' English language proficiency. The course covers the four language skills (both receptive and productive). That is, listening, speaking, reading and writing. Students have to sit for an English language placement test to determine their English language proficiency levels. During their first academic year, students are also trained in computer skills, and are required to successfully complete a set of modules in computing. However, the computer skills module is not assessed.

According to the statistics provided by the MOHE, a total number of 2010 students were admitted to the foundation courses which were organized and conducted by the six Omani colleges in the 2005-06 academic year. Hence, the importance of the foundation year to the students that no student is allowed to attend academic courses in his/her discipline area until they have passed the final test in the general English language skills given to them at the end of the two semesters.

At the beginning, a diagnostic test is conducted to assess all first year students. The aim of this test is to determine their level of English, and accordingly decide on the nature and level of the language course they should attend. After attending the designated courses, they are required to sit a mid-and an end-of-term achievement tests in order to monitor and assess their improvement during and at the end of the first and second semesters.

At the end of the first (or foundation) year, a final exam is carried out to assess students' achievement and determine whether or not they have met the requirements which are set by the MOHE in cooperation with the Hawthorn Centre (a specialized English language teaching centre) which is an institute that is overseen by the University of Melbourne in Australia. The centre's role consist of preparing the exam with the MOHE supervision, supervising the marking process and providing the Omani Ministry of Higher Education with the exam's results as well as statistics on students performance and the level they have managed to achieve. In order to pass the exam, each student has to get an overall mark which is equivalent to band 4.5 in the IELTS (International English Language Test System) otherwise they would not move on to the specialized subjects which they wish to do in their college.

So, each and every student in the first year has to go through three different types of tests; first, the diagnostic test which is aimed to identify or determine the students' level; second, the achievement test which is intended to get academics to find out or assess the students' progress throughout the first year; and third, the final exam which also aims to check on the students' proficiency level and the English language skills they have acquired during the foundation year All of these tests are carried out in the traditional paper-and-pencil assessment mode.

However, although colleges and universities require all the students to take the foundation course, those students whose English language proficiency level is high, and can prove this by presenting an IELTS or a TOEFL certificate, are exempted from taking the foundation course, which means that they can proceed directly to their academic study subjects.

3.4. The History of Computer Use in the Sultanate of Oman

Right from the outset, we should be clear that Omanis have a relatively short history in computer use. Until the last decade, the majority of Omanis had known almost nothing about computers. In addition, computers had been rarely used in both the government and the private sectors except for some companies or institutions, largely banks. Gradually, computers started to invade both the government as well as the private sectors.

As far as education is concerned, both schools' administrative staff and teachers had no idea about what a computer might be and how much effort and time it could save. Consequently, students were also unacquainted with the computer, for this machine was never accessible to them. However, in the recent years, things have completely changed. It can easily be noticed that now Omani people have access to the computer almost all over the country and that computers have become widespread in cities and towns including some small and remote villages. In fact, computers are not only used in offices and schools today, but also at homes.

However, despite all these cultural changes, there are still many students who had never dealt with a computer until they reached university. This might be attributed to various reasons. In some cases, the student might have come from a remote village which had no electricity supply. Other cases might include students who have not been acquainted with computers at school or at home. In fact, whatever the reason might be, we have to admit the fact that the student population in Oman is quite varied, so we have to take into consideration students' experiences when approaching the history of computer use in Oman.

In terms of technology, there has been a noticeable growth in the use of the internet in the Sultanate of Oman in the recent years. The internet users in 2002 were 90,000 (3.9% of population) while in 2008 this rate rose to 300,000 (9.1% of population) (http://www.internetworldstats.com/me/om.htm). Despite this increase in the rate of internet users in the Sultanate, it remains relatively modest; particularly when it is compared to the percentage of internet users in some developed countries like the UK, where in 2002, 45% of householders had access to the internet at home and in 2009 this rose to 70%. Besides, about 96% of British aged 16-24 had regular access to the World Wide Web according to a recent report released by the Office of National Statistics ('http://www.statistics.gov.uk/stat Base/ssdata.set.asp'). Unfortunately, there is no similar source of information to date to reveal the number or percentage of people having computers at home in Oman.

Another major difference between both countries relates to cultural factors. Unlike girls in the UK, females in Oman cannot even today go to internet cafés as it is considered culturally and socially odd for a female to enter such places. However, males have the freedom to frequent internet cafés and they often do. Cyber cafés are widespread in the capital, Muscat, and most big cities and even in some small towns we could find a few such cafés. So, the only possible way for an Omani female to use the internet is either at home (if the family has a computer) or at school or university.

3.4.1. Computers use in Omani schools

In the Omani educational system, computing only started to be taught as a school subject in the 1998 –99 academic year. That is, along with the Basic Education system. As noted earlier, this first included only seventeen schools from different regions of Oman. According to the statistics provided by the Information Technology Department in the Ministry of Education, the schools in which computing is taught were increasing every year, reaching 221 cycle one Basic Education schools and 132 schools having both cycle one and cycle two in the 2003–04 academic year (Ministry of Education, 2006). Each cycle 1 school (grade 1-4) in the Sultanate is equipped with new Learning Resource Centres (LRCs), each comprising 12 to 14 computers. As for cycle 2 schools (grades 5-10), each was equipped with two computer labs, each comprising 30 to 35 computers, in addition to a learning resource centre. Consequently, new computer teachers as well as computers lab coordinators have been recruited. Therefore, teachers today have access to those computers, and are encouraged to use them to print out their worksheets, lesson plans and also their tests.

The substantial change that has occurred consists in the step which has been taken by the Ministry of Education. This step consists in the teaching of computing which is now considered as a basic school subject New syllabuses have been devised to facilitate the teaching of this new subject Teachers have also been supplied with the needed materials and course books to teach such an important subject.

Students attend computing lessons which are meant to enable them to successfully use the computer and acquaint them with basic software programs like Word, Excel and Power Point. Students are also encouraged to carry out projects and create electronic portfolios on which they are assessed according to the practicality of their work and how genuine and creative their ideas are. Furthermore, students' contests are held throughout the Sultanate schools aiming to encourage students to design websites as well as other educational games or programs. Much effort has also been made to get students to take part in the e-school events. Schools are now connected to the internet, using specially designed software programs for school–related purposes.

Fortunately, schools in Oman are now supporting computer use. In fact, the whole school environment encourages the use of information communication technology (ICT) to the extent that students can now get their school results by simply logging on to the Ministry Of Education's website.

However, as computer Literacy only started to be taught as a subject in Omani schools from the 1998/1999 academic year, this will lead us to assume that Omani students will have similar background in computer experience when they reach the university in 2010/2011 academic year, since they are still studying at schools. And as they have been exposed to computers since grade one; it might affect their attitudes towards computers. However, my study sample was not from these students, so will that have an effect on students' performance on CAA, and if so what the effect of computer experience and student attitudes towards CAA performance would be, and that what this study will try to explore.

3.4.2. Computer Use in the Omani Higher Education Institutions

Today, colleges and universities in Oman have witnessed a noticeable change towards technology. The classrooms are now well equipped with computers, slide projectors and video projectors which enable tutors and lecturers to use the Power Point program. In addition, the MOHE has recommended that computers be used in all Omani colleges and universities, and in the private ones. As a result, computers are now used not only in classrooms but also in libraries, lecture halls and laboratories which are also connected to the internet

In addition, the MOHE started in 2006 an electronic project in cooperation with the MOE to help students to register for college or university courses. They have established centres (Higher Education Admissions Center) where students, who have finished grade 12 and managed to attain a certain score, can go to apply for electronic registration in any college or university without having to queue or do lots of paper work and formalities The application for enrolment is usually confirmed within a few days. This process has not only facilitated students' registration but also made it easier for them to opt for the most suitable universities and study areas or disciplines for them. So, instead of having to submit all their documents to one university and possibly lose the chance of getting other ones each student is now guided and helped to make the right choice. So, today technology and computers are regarded as the backbone of education in Oman, as in many other countries.

Although Computer Assisted Assessment has not been established yet in Oman, there are some institutions which are already using computers to enhance the teaching and learning process, and this innovative teaching method has already been faced with lots of resistance and even rejection from some stakeholders. This could be due to various reasons. There are a number of research studies which have recently been carried out in Sultan Qaboos University (SQU) and which reveal that academics' reluctance to use technology is caused by their resistance to change (Akinyemi & Al-Musawi, 2002; Al-Saleem, 2006).

Akinyemi & Al-Musawi (2002) state that this resistance may be due to any of the following factors:

- Fear of redundancy. That is, faculty members replacement by technology.
- Complacency of the faculty members about set forms of practice.
- Negative beliefs of the faculty members towards using computer technology in teaching and learning.

Al-Saleem (2006) notes that no attempt has so far been made in Oman to explore the faculty members' beliefs about information communication technology (ICT) or to investigate how they (faculty members) make sense of their professional realities, how they influence their classroom practice, or how they mediate upon the interpretation of their teaching tools. Moreover, it is argued by Al-Saleem (2006) that faculty members have been resistant to change due a number of reasons such as heavy working load, lack of technical support from their colleagues and also having negative attitudes towards computers. The other factor which has strengthened this resistance to change is the lack of pressure from the university administration (Al-Saleem, 2006).

EL-Shibiny (1995), cited in Al-Rabiey (2002, p.62) summarizes the challenges facing the application of modern technology in educational institutions in Oman as follows:

- The lack of identification of the kind of technology most suitable to realize or achieve the highest degree of advancement of the Omani society.
- The need to prepare national professional manpower of the highest quality to bear the responsibility of integrating technology in the Omani communities.
- The need to constantly keep up with the rapid changes and the numerous innovations in technological inventions.

As mentioned by A'Sadoun (1998) there are some research and empirical studies have revealed that the computer has many features which make it a good educational mediator, but this depends on the availability of suitable software programs. Furthermore, training teachers is also needed so that they could use it in an easy way, which facilitates their work and helps them to meet the purposes sought (A'Sadoun, 1998). Nevertheless, these studies admit that there are many obstacles that educationalists' face when using technology for educational purposes, and note that these obstacles prevent them from using the computer efficiently in educational institutions. For example, Sa'ada & Saratawi (2003) mentioned some of those obstacles which we would list below:

1. Shortage of professionals working with computers in the educational field in many countries.
2. The limited number of computer users and the lack of awareness of the importance of using computers for educational purposes, especially in developing countries.
3. The unavailability of sophisticated, proper and suitable software programmes, which is due to the big effort needed for writing and designing them.

4. Using computers in education is, to some extent, very expensive.

In addition, the e -learning software that has been adopted and applied in Sultan Qaboos University has increased students' enthusiasm to learn and study with the help of computers. The results of the study carried out by Al-Hanaie, (2005) have also revealed that using the computer and multimedia has increased students' motivation to learn and as a result, when assessed, they (the students) could achieve better results.

We could assume, then, that this software program has encouraged them to learn, which has positively affected their results. It is worth mentioning that the computer program which was applied in SQU provides students and academics with instant or timely feedback. Al-Hanaie (2005) has also noticed the positive effect of timely feedback on low-achieving students while carrying out the experiment. In fact, instant feedback boosted student' motivation and fostered their learning by encouraging them to continue when their answers were correct and to try again when they got wrong answers, and this confirms with the theory which asserts that timely feedback is constructive as it reinforces the learning process (Al-Hanaie, 2005).

As for the situation in Oman, there is still much to be done so that higher education institutions develop further. However, we can summarise the main difficulties which Omani higher education institutions face as follows:

1. Most academics working for the Ministry Of Higher Education institutions in Oman do not have sufficient technological background and refuse to use technology for educational purposes, especially in teaching.
2. There is a need to change academics' attitudes towards computers and get them to develop more positive ones towards technology.

3 There is a need to identify the different types of obstacles which could be faced with using technology.

4. There is a pressing need to well-prepare and train teachers to be able to use technology so that they would be able cope with any development in the educational system in the future.

Although the Omani Ministry Of Higher Education is a relatively young institution when compared to ministries of higher education in other countries, and although Omani higher education institutions are well equipped with sophisticated technological equipment, there is still one problem in these institutions that should be overcome, for academics are not applying this technology the way it should be. That is, academics are reluctant to use technology for educational purposes, so we might assume that one of the obstacles which prevents the use of technology in Omani higher education institutions is academics' clinging to the traditional teaching and assessment methods.

Finally depending on the previous research about Computer Assisted Learning (CAL) which has been conducted in Oman, we might wonder about academics' reaction and attitudes towards CAA So, the question that should be answered is whether or not their attitudes towards CAA would be the same as their attitudes towards CAL, especially as CAA would diminish the work load they have and reduce the heavy burden they are carrying by having to assess a rapidly increasing student population. Hence, what this research would try to achieve is to explore students' and academics' attitudes in Omani higher education institutions towards CAA to answer that question.

3.5. Student Assessment in Oman

Because this research project will focus on student assessment by computer in Oman, it is important to put the study in the context of current assessment practice. It has been mentioned earlier that before 1970 the educational system in Oman was equated with 'Quran schools' where both teaching and assessment were carried out by 'sheiks' or 'Imams'. As a result, learning was mainly based on rote-learning or memorization. Hence, assessment according to this system of learning was through oral examinations (Al-Alawi, 1997). The assessment of students used to be carried out by the 'Sheik' or the 'Imam' whose evaluation was trusted by society and met by wide acceptance. After 1970 when modern schools were established, assessment moved towards more quantitative measurement based on paper-and pencil testing. Paper–based testing has become the most popular method in assessing student learning (Palomba & Banta, 1999).

However, neither formative nor continuous assessment used to receive any attention. It was assumed, then, that quantitative measurement was the most reliable and objective method to judge or evaluate student learning. Hence, most of the teaching and learning activities used to revolve around memorizing facts just to pass exams.

Broadfoot (2000) cited by Al-Alawi (2004) argues that assessment is part of the culture of the people and that all the tools that are created to meet the need of a particular time and place are based on the assumption that there is only one way of solving the problem. Therefore, it will be very difficult to escape from the power of the traditional method.

In Oman, norm-referenced assessment used to be quite prevalent. It was assumed that the assessment which is based upon a particular norm is fair, valid and reliable. However, in recent years, norm-referenced assessment has been criticized because it only focuses on classifying students and discriminating between them based on a normal distribution of score. Therefore, tests were set according to how well they would rank students and discriminate between high and low achievers, and not according to pre-set criteria that are meant to identify students' needs in order to improve their learning by improving the applied teaching activities (Al-Alawi, 2004). Murphy (2001) argues that this assessment method (Norm-Referenced Testing) will disadvantage a cohort of students and favour another because they can perform better or get a higher score in an exam than the other low-achieving students.

In contrast to norm–referenced, criterion-referenced tests are meant to get students to demonstrate what they can do. That is, this type of assessment is aimed to get students to accomplish what they are expected to in a way which focuses on what they know rather than on how they compare with the others

However, in Oman, summative assessment used to receive full attention whereby interest was in achievement tests which usually take place at the end of a course of instruction. Things have recently completely changed, for focus is now laid on formative and continuous assessment which takes place during the learning process and is meant to help teachers to spot the problems which might be faced by their students and remedy them while there is still time for change.

Hence assessment is now viewed as a process that is integrated in the teaching activity, and as a tool which is rather meant to enhance student learning than make judgments about students. So there is a demand on new method which can enhance student learning efficiency.

There is no doubt that changing culture is not an easy process as culture is related to both history and geography. That is, it represents what a given society has inherited and acquired over time within a well-defined geographical context. The most important aspects of culture relate to values, beliefs and perceptions which might be invisible at times but very effective indeed.

It is assumed that the traditional assessment methods rely on a single-measure test score, which means that the assessment culture is facing many challenges today as it does not really assess student learning or lead to learner autonomy (Broadfoot, 2001; Simpson & Hayward, 1998).

Black (2000) also notes that "the traditional assessment culture that is based on written tests reduces teacher and student autonomy and leads to students being taught how to pass the exam" (p.412). What could also be inferred from the last definition is that this type of assessment (traditional assessment) limits rather than enhances student learning.

Moreover, this assessment type is based on the assumption that the best way to get students to learn better or make progress is by comparing their performance with the performance of their peers rather than by comparing it with well-defined and pre-set criteria. So, as noted by Black (2000) assessment becomes "based on mark, grades [which] emphasize competition rather than personal improvement" (p. 409).

Changing the culture of assessment in the Omani society is considered as one of the most difficult missions, especially when we consider the many challenges brought about by either educationalists or by the students and the assessment culture and its development in Oman.

Actually, we should relate those challenges to the pre 1970 era as previously presented. At that time, assessment was manly dependent on one person (Imam or Sheik) who used to teach the students the main teachings of the religion. Therefore, the assessment of learning was focused only the students' capacity to learn by rote or memorise facts. Assessment in those days was not formal or systematic Hence, all the students used to be taught together, at the same time and regardless of their age or gender. As for teaching time, it used to last for three to four hours daily. Later, the assessment principles started to develop gradually with the establishment of the formal and systematic education which started to cover all the regions of the country. However, even then assessment was carried out solely for summative purposes. That is, assessment was commonly used to upgrade students from one level to another.

Soon, the Omani society started to accept the new educational system and embrace its way of assessment. People also started to understand its marking and grading system gradually This new system persisted for many years and was used to differentiate between the students' different skills and compare their performances with their peers'. So, parents started to feel proud of their children's performance just because they got better grades than the others, which was the most important thing for them. However, little attention was given to students' low academic achievement. The culture of summative assessment had continued from 1970 till lately and had become part of the assessment culture, for it was considered as the best assessment method.

Nevertheless, starting from the late 1990s, the notion of continuous assessment started to call educationalists' attention, which later resulted in it being considered and applied alongside summative assessment. Formative assessment has recently become a reality and has been amalgamated into the Basic Education system. However, although it is now around nine or ten years since formative assessment was first used in Basic Education schools, there are still some people who are attached to the summative assessment system and refuse to recognize the benefits of formative assessment. This simply indicates that bringing about any kind of change in the current assessment system will take time before it is accepted by the Omani society or becomes part of the assessment culture in Oman. Hence, any innovative assessment method is foreseen to encounter many challenges created by the traditional assessment culture.

However, we should note that by speaking about new assessment, we do not mean discarding traditional assessment methods which depend upon summative testing, but rather making use of this assessment type along with other assessment types to meet other assessment purposes which include: diagnostic and continuous or formative.

In this research context, the focus would be laid on Computer Assisted Assessment as a new assessment mode and to ensure the equivalency between paper and computer mode of assessment. CAA importance comes from the fact that it can be used for diagnostic, formative or summative assessment, so, this research could be viewed as an attempt to elevate the level of student learning rather than just measure it.

3.6. Summary

Undoubtedly, there is a strong link between assessment and culture So, in order to understand any assessment system, we should understand the Socio-cultural context in which that system is applied. Moreover, changing the assessment culture of society is not easy so, it has to be carefully dealt with, for any innovative method would first be faced with resistance and even rejection unless its benefits to students and staff are clarified and pointed out.

Chapter Four: Computer Assisted Assessment Equity and Score Comparability

4.1. Introduction

Changing the assessment method from paper test to computerised one involves considering the equity of test scores. The equating of test score is complicated process which "involve small statistical adjustment to account for minor differences in the difficulty and statistical properties of the alternate forms" (American Educational Research Association American Psychological Association & National Council on Measurement in Education, 1999, p.51). When the statistical properties are met between computer and paper test, computerised testing will be used for students with equity and fairness to some extent. Therefore, in the comparability studies, many factors such as the effect of computer experience, computer self efficacy and attitudes towards computerised assessment on students performance should be examined to ensure the comparability between paper and computerised testing.

This chapter will comprise the literature review of comparability studies which will be composed of two parts. Part one will be devoted to summarizing the research studies which compare the two testing modes, computer-based tests (CBTs) and paper-based tests (PBTs). Focus will also be laid on the impact of such variables as: gender, computer experience, computer self-efficacy, staff and students' attitudes towards computers, and other demographic factors

Part two will evaluate the findings of the research studies cited in part one, the methods which were applied as well as the results obtained and the conclusion.

Before presenting the findings of the literature review, it should be noted that score comparability or equivalence between the two assessment modes CBTs and PBTs is defined by the American Psychological Association [APA] as follows: "Scores from conventional and computer administrations may be considered equivalent when (a) the rank orders of scores of individuals tested in alternative modes closely approximate each other, and (b) the means, dispersions and shapes of the score distributions are approximately the same, or have been made approximately the same by rescaling the scores from the computer mode" (American Psychological Association, 1986, p.8). The Guidelines also stress the need to limit the impact of such variables as computer anxiety, and computer experience for validity reasons.

The articles presented in this literature review fall under the following headings

- There is no difference: This part will include the research studies which have concluded that there is no substantial difference between both modes of administration.
- There is a difference in favour of computer-based assessment: This part will comprise the research studies which have concluded that there is a significant difference between both assessment modes and that computer-based tests are much 'better' in some respect, such as they provide room for the inclusion of better quality written comments, and offer a higher degree of reliability and equity.
- There is a difference in favour of paper-based assessment: This part will be devoted to presenting the studies which have concluded that there are slight differences in favour of paper-based tests.

- This part will be aimed to present and discuss the potential impacts of such variables as computer experience (familiarity), student and staff members' attitudes towards CAA as well as the possible effects of computer anxiety, gender and socio-economic status on the examinees taking computerized tests in the light of the published literature on the comparability of assessment modes.

4.2. Studies of Differences in Relation to the Mode of Assessment

The subsequent sections present the studies with different finding followed by literature review evaluation.

4.2.1. No Mode Difference Found

When students are not affected by the delivery mode that means their performance is determined by the tests' content only. There is much in the literature to suggest that there is no difference which might be brought about by changing mode of delivery as many studies have traced no difference between students' achievement scores in both delivery modes, CBT and PBT across a wide range of tests.

Katz & Dalby (1981) report having found no difference in the Eysenck Personality Inventory Test scores across both delivery modes. Holden & Hickman (1987) have also come to the conclusion that there are no significant differences between the scores obtained by students when taking both computer-based and paper-based tests. Similarly, Wang, Young & Brooks (2004) have come to the same conclusion when discussing the results of an administration mode comparability study which was conducted using the Stanford Diagnostic Reading Test (SDRT4) 4th Ed. and the Stanford Diagnostic Mathematics Test (SDMT4) 4th Ed. which were delivered in both computer-based and paper-based modes. Wang, Young & Brooks (2004) note that "the results of the study provide strong, broad-based evidence of the reliability and comparability of SDRT4 and SDMT4 scores for all grades and levels regardless of the administration mode" (p.3). Citing (Bergstrom, 1992; Boo & Vispoel, 1998; Bugbee Jr, 1996; Evans, Tannehill & Martin, 1995, Neuman & Baydoun, 1998; Wang, Newman & Witt, 2000) Wang, Young & Brooks (2004, p.3) point out that "studies conducted by other researchers typically find that measures obtained from tests delivered in the computer based mode are similar to those obtained from the paper-and-pencil mode ".

Brown & Augustine (2000) have traced no significant differences across testing modes. As they concluded "This study revealed no significant differences between the performance of students completing the pencil and paper format version versus the screen reading format when controlling for reading performance"(p.23).

Akdemir & Oguz (2008) aimed to compare Turkish undergraduate students between paper and computer based test They found no significant mode differences between students performance. Also there were no mode differences between males and females. They concluded that "This study has showed that student achievements do not vary on the administration of computer-based tests and paper-based tests which indicated that computer-based testing could be an alternative to paper-based testing for Turkish students" (p. 1203).

After examining the potential impacts of changing the presentation form of a reading comprehension test (from paper-and-pencil to computer-based) on 4th grade students' test scores, Higgins, Russell & Hoffmann have concluded that there is no noticeable mode effect and stress that "there were no significant differences in reading comprehension scores across testing modes" (Higgins, Russel & Hoffmann, 2005, p.30).

Similarly, after assessing the impact of the mode of delivery on a student cohort of 644 seventh graders in mathematics Poggio et al., (2005) have deduced that there are no significant differences between the scores which the students had obtained in both assessment modes. They note that "descriptively, there was a very little difference in performance (less than one percentage point). Results make clear that there existed no meaningful statistical differences in the composite test scores attained by the same students on a computerized fixed form assessment and an equated form of that assessment when taken in a traditional paper and pencil format" (Poggio et al., 2005, p.3).

Zandvliet & Farragher's comparative study (1997) stresses that although students may have a preference for one particular testing mode, empirical evidence may reveal no actual difference across delivery mode. They point out that "the results of the comparative analysis of test scores indicated that there were no significant differences between computer-based and written test scores [although] survey responses indicated a student preference for the computer-based test over the written format" (Zandvliet & Farragher, 1997, p.423).

This finding also was supported by Texas Educational Agency (TEA), 2008 which conducted a review of literature and found that 43 out of 64 comparability studies, which examined different content areas such as mathematics, English language, arts, science and social studies, found a comparable result between overall test score in paper and computer test. However, 12 studies found that computer was a harder mode and 9 studies found the paper mode is harder (Texas Education Agency, 2008). Also Wang *et al.* (2007, 2008) conducted two

Meta analysis studies and found a comparable result between two modes in student mathematics and reading scores

It is obvious enough from the findings of the studies cited above that there is much in the literature which supports the idea of equivalence between the traditional Paper Based Testing (PBT) assessment mode and the Computer-Based Testing (CBT) mode, particularly in terms of students' achievement scores.

4.2.2. Mode Difference: Computers are 'better'

As a result of the substantial development and advancement in computer technology and the noticeable increase in computer use worldwide, numbers of studies have concluded that students attain better scores when they take Computer-Based Assessment than when they take the traditional Paper-and-Pencil one.

Austin & Mahlman (2000), for instance, have compared and evaluated the scores which students attained in vocational competency after taking an administrative office technology test in both assessment modes, computer-based and paper-based testing. They have come to the conclusion that "the direct comparison [..] indicated that the internet test scores were higher than the paper and pencil version" (p. 11). Empirical evidence also includes the results of an attitude survey which has revealed that students' reaction towards the use of the internet was very positive Austin & Mahlman note that "the attitude survey responses indicated very positive evaluations of the process of internet testing and, correspondingly positive reactions to the overall experience and performances for the internet format" (2000, p.11). Austin & Mahlman (2000, p.1) also point out that "a majority of students responded positively to global evaluations (75% reported a positive experience; 62% preferred or greatly preferred internet format over the traditional)". It is worth noting that Austin & Mahlman's study included not only students' responses but also teachers'. Teachers, too, responded positively to conducting vocational assessment on internet. It is noted that "qualitatively, teachers/facilitators reported that they and the students were impressed with the rapidity of reporting of results via an overall score on six scoring clusters (ranging from office equipment/procedures to professionalism)" (Austin & Mahlman, 2000, p.11). Hence, they have come to the conclusion that online assessment (despite the few problems encountered such as: accessing the targeted internet site, downloading and submitting the test has more advantages than the traditional Paper-and-Pencil assessment mode. They note that "results suggest that the advantages of testing via the Internet outweigh the disadvantages (Austin & Mahlman, 2000, p. 11).

Similarly, Kapes & Vansickle (1992) also found that the computer-based administration mode is more reliable than the paper-based administration mode. This means that CBTs can meet the various aims of assessment which are reliability and equity.

While investigating the impact of administration modes on the students taking open-end test items taken from the Massachusetts Comprehensive Assessment System (MCAS) and the National Assessment of Educational Progress (NAEP), Russell (1999) reveals having traced some statistical differences among grade 8 students' scores in the science test across administration modes, and notes that "for the science test, performance on computer had a positive group effect" (Russell, 1999, p.1).

Similarly, investigating the impact of delivery mode, Russell & Haney (1997) reveal that the students who are used to writing on the computer perform better in computer-based testing. They point out that "the findings show that [...] for students accustomed to writing on computer, responses written on computer are substantially higher than those written by hand (effect size of 0.9 and relative success rates of 67% versus 30%)" (Russell & Haney, 1997, p 1) It is large difference between the two modes; however, it is surprising that there were *no mode differences in the multiple choice question subtests* which conducted in the same study which made it hard to explain.

This suggests that adopting computer-based rather than paper-based tests would yield better student performance scores, particularly in the light of the rapidly increasing computer familiarity rates Russell & Haney (1997) also emphasize the fact that the substantial growth of computer use at schools and among students would certainly have various implications, for denying the students who have a high computer familiarity level their right to opt for sitting for computerized tests would give rise to validity and equity problems. That is, *those accustomed to writing on the computer would be disadvantaged if made* to take paper-based assessment. Russell & Haney assert that "as increasing numbers of students grow accustomed to writing on computers; these [paper] assessments may yield underestimates of students' writing abilities" (1997, p.1).

Ricketts & Wilks (2002a) have investigated the effect of introducing Computer-Aided Assessment in some 1st and 2nd year modules at the University of Plymouth. Their evaluation shows that although Geography students were not motivated with the introduction of CAA, Biology and Business students have shown a great acceptance of the new assessment mode. Ricketts & Wilks note that "only 55% of students in Geography preferred on-line examinations compared with 72% to 90% in the other subject areas" (2002a, p.307).

Dolan et al (2005) also sought to compare the effect of both administration modes on students' performance in the reading comprehension of short and long passages. The findings have revealed differences in students' performance in dealing with short and long passages across administration modes. Students' scores were higher in the computer-based test of long reading passages (mean=76.6%) than in the paper-based mode (mean= 55%). This significant difference contrasts with students' scores in dealing with short passages for which students got similar scores in the paper-based testing mode (mean=60%) to the computer-based testing mode (58%), with no significant statistical difference. Moreover, qualitative results have supported the use of computer-based testing as students preferred this assessment mode to the traditional paper-based one.

4.2.3. Mode Difference: Paper-and Pencil is 'better'

In contrast to the research studies cited above, there are several other studies which have displayed some evidence supporting the paper-and-pencil assessment mode.

While attempting to estimate the amount of change in students' proficiency level after receiving an enrichment program in Additional Mathematics at Ngee Ann Polytechnic, Singapore, Lam & Tham-Ng (1996) have come to the conclusion that the testing mode certainly affects students' performance, and they have reported that students get better scores when they are assessed in the traditional paper-and-pencil mode. They point out that "there are evidences to show [...] that test mode does affect performances favouring paper-and-pencil to computer mode" (p 7). However, Lam & Tham-Ng note that experience with computerized testing may change the mode effect. They note that "experience with test on computer possibly took care of the computer mode disadvantages" (Lam & Tham-Ng, 1996, p.1).

Investigating score comparability, Choi & Tinkler (2002) examined grade three and grade ten students' performance in math and reading across both administration modes. They have found that computer-based reading tests were more difficult than the paper-based ones especially for 3rd graders. However, they found no significant differences in the mean item difficulties for 10th graders in the math test. Choi & Tinkler relate this to experience with computers as they have found that computer test performance was related to computer familiarity That is, students with a lower computer familiarity level tended not to perform as well as those having a high level of computer familiarity when taking mathematics and reading tests administered in the computer-based mode. However, they suggested that acquiring more computer familiarity would result in eliminating the test mode effect. Choi & Tinkler (2002) point out that "taking tests online seems to present a greater novelty effect to young students than to older students [and] more frequent exposure to such online exams may eventually eliminate the novelty and mode effect" (p.10). They also point out that as the reading passage (especially for long reading passages) can not fit onto a single computer screen, this could make it more difficult for students to read from the computer than to read from a hard copy (Choi & Tinkler, 2002).

This problem was also stressed by Ricketts & Wilks (2002a) who note that although their study has shown that students prefer computer-based assessment for immediacy of scoring and providing timely feedback, they nevertheless may find it difficult to read from a computer screen. They write "we have already reported that the change of screen presentation between 2000 and 2001 in the biology module produced a large improvement in student acceptability of computer-assisted assessment" (p.310).

However, Choi & Tinkler (2002) and Bridgeman, Lennon & Jackenthal (2003) cited in Peak (2005) note that "other studies have indicated significant mode effects when students must scroll or navigate through information on the computer screen to answer test questions [and that] these findings have not been universally accepted. For example Pomplun, Frey & Becker (2002) asserted that difficulties in reading on computer screen were related to primitive technology" (Peak, 2005, p.9).

Hence, we might conclude that with advances in computer technology, this test mode effect would finally vanish and students would feel more comfortable taking reading tests on computer. Peak (2005, p.17) notes that "the general findings of comparability across modes may be due to the fact that modern computer test systems allow students to navigate the computer tests as they would a paper test, thus allowing for similar test-taking strategies across modes".

Lee & Weerakoon (2001) have also revealed that while investigating the role of computer-aided assessment in health professional education, it has been revealed that students' performance was significantly better in the paper-and-pencil test than in the computer-based test. They point out that "overall, the cohort performed significantly better in the paper test than in the computer test, with an average total mark 75% higher and average mark for common questions 3% higher in the paper test" (Lee & Weerakoon, 2001, p.156).

A few other studies have come to the same conclusion (Cerillo & Davis, 2004; O'Malley et al., 2005). Peak (2005, p.14) notes that "Cerillo & Davis (2004) discovered that students performed 4-7% better on the paper version of a high school graduation test compared with a computerized version". However, Peak attributes this to a limitation in the study, for according to Peak the students who had taken the computerized test were not motivated enough and lacked the incentive as this test was not meant to count towards graduation. Peak (2005, p.14) asserts that "unfortunately, in this study, only the paper test counted towards graduation, so the incentive to perform on the computer tests was not similar to that on the paper tests".

Bunderson et al., (1989) conducted a review of paper-based and computer administrated tests. They found nine studies that showed superiority of paper based testing. They assert "...the score on test administrated on paper were more often higher than on computer administrated tests....the score differences were generally small" (p.378).

4.3. The Effects of Computer Experience, Computer Self-Efficacy, Demographic Factors, and Attitudes towards Computer Assisted Assessment

4.3.1 Computer Experience (Familiarity)

In most of the studies which have addressed and examined the comfort level that examinees have with computers, focus has always been laid on the notion of familiarity with computers and how it might affect the examinees' performance when they take a computerized test. For accuracy, we should note that this notion includes "the familiarity of test-takers with the computer itself

[...], and familiarity with the manipulation of a mouse with two buttons [and] if the test is delivered over the internet using a standard browser such as Netscape or internet explorer, familiarity with the WWW should be taken into account" (Fulcher, 1999, p.292). In some cases, it has been found that those examinees that are less familiar with computers do not perform as well as those who are more familiar with computer use when they take a computerized test. In such cases, test-designers may be inclined to administer the test on paper in order not to disadvantage those examinees that are less familiar or may be unfamiliar with computers use (Russell, Goldberg & O'connor, 2003, p.8).

The bias which could be brought by the use of CBTs was pointed out by Fulcher (1999) who investigated the potential bias in a computer-based English placement test in relation to computer familiarity and accordingly to students' attitudes towards taking CBTs. According to Fulcher (1999) lack of computer familiarity may affect students' scores on computerised test. This study also stresses the fact that it is not only familiarity with the testing mode which is required (however important it is), but also familiarity with computers so that a test accurately measures what it purports to measure without being affected by any compounding variables, which might give rise to validity problems.

Bunderson et al., (1989), cited in Fulcher (1999, p.291) conclude that "lack of familiarity with computers could be assumed to be a major factor in achieving lower scores on computer-based tests". Hence, lack of familiarity with computers would cause equity issues, for students' achievement might be affected by other variables than the test's content.

Russell (1999) also stresses that computer experience (prior computer usage rate and keyboarding speed) could tremendously affect students' performance either positively or negatively when they take CBT. He points out that students' performance could be impeded by low keyboarding speed. However, he asserts that this impeding effect would soon vanish with increased familiarity with computers. He notes that "this effect became less pronounced as keyboarding speed increased" (Russell, 1999, p.1).

4.3.2. Computer Self-Efficacy

Several studies have addressed the notion of self-efficacy and investigated the potential impacts it might have on students' performance. First used by Bandura (1977), the term *self-efficacy* relates to an individual's perception of the *extent to which he/she can execute a course of action to meet the demands of a* certain situation. Several definitions of self-efficacy have been developed ever since this concept was introduced. Busch (1995), for example, defines it as "the *belief in one's ability to execute successfully a certain course of behavior*" (p.147).

More comprehensive definition is provided also by Bandura who notes that perceived self-efficacy refers to "people's judgments of their capabilities to organize and execute courses of action required to attain designated types of performances. It is concerned not only with the skills one has but also with judgments of what one can do with whatever skills one possesses" (Bandura, 1986, p.391).

What could be inferred from the last definition is that self-efficacy relates to the psychological construct of an individual concerning their capability of performing a particular thing successfully or attaining a particular objective It is also understood from the definition that this concept is context-specific or 'domain sensitive' as noted by Kurbanoglu (2003). That is, it relates to self perception and self evaluation in a specifically defined situation. Kurbanoglu (2003) notes that "because self-efficacy is based on self perceptions regarding particular behaviour, the construct is considered to be situation-specific or domain sensitive. That is, an individual may exhibit high levels of self-efficacy within one domain while exhibiting low levels within other domains" (p.636).

According to Bandura (1977), there are four sources of information which affect the attainment of self-efficacy: performance accomplishments, vicarious experience, verbal persuasion, and emotional states. While performance accomplishments relate to the individual's own experiences and their success or failure in accomplishing a particular task, vicarious experiences are rather related to watching or reading about other people's accomplishments. That is, observing others succeed or fail. As for verbal persuasion, it is concerned with such speech acts as encouragement or appraisals. The forth source of information includes such emotional variables as: anxiety, apprehension or stress. According to Bandura, the most important or influential source is performance accomplishments because any individual will first build on his/her own experiences before considering the others (Bandura, 1977).

It should be noted that our focus in this study will be laid on one particular type of self-efficacy which is computer self-efficacy, for it is relevant to our research as "researchers to date confirm that computer self-efficacy not only determines decisions by individuals to accept and use the computer system, but is also a good predictor of achievement in computer-related tasks" (Torkzadeh, Koufteros, & Plughoeft, 2003, p. 264).

Computer self-efficacy is defined by Compeau & Higgins (1995, p.192) as "a judgment of one's capability to use a computer". A similar, though broader, definition is the one provided by Stephen & Shotick (2002) in which computer self-efficacy is defined as "an individual's belief in their ability to use technology in order to solve problems, make decisions and to gather and disseminate information"(cited in Stephens, 2006, p.29). Hence, we might assume that computer self-efficacy relates to an individual's attitudes or perceptions of himself/herself in the use of computer technologies (Delcourt & Kinzie, 1993).

According to a study conducted by Wallace (1999), there are four main factors which can influence the development of computer self-efficacy. The four factors are: *computer anxiety, computer confidence, computer liking, and computer* knowledge. However, it is obvious that the four factors postulated by Wallace are inherently interrelated as computer knowledge, for example, might contribute in reducing computer anxiety, which in turn increases computer confidence levels and results in increasing computer liking. Similarly, computer liking accounts for seeking to acquire computer knowledge which might lead to reducing computer anxiety and increasing computer confidence. The diagram below (Figure 4) gives an illustration of the four factors' interrelatedness.

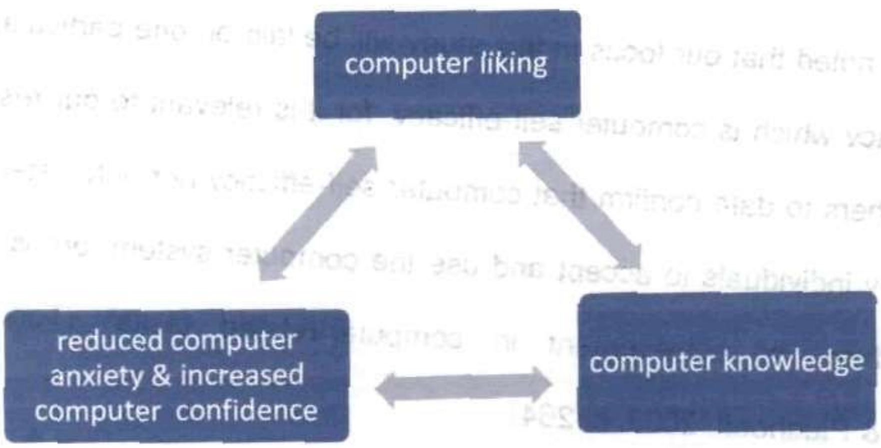


Figure 4. Factors Affecting Computer Self Efficacy

As computer knowledge is generally determined by computer experience, we might then come to the conclusion that computer experience is a considerable determinant of computer self-efficacy. Actually, several studies have concluded that the amount of computer experience that an individual possesses usually has a positive impact on that individual's attainment of self-efficacy (Busch, 1995; Johnson, Ferguson & Lester, 1999; Moroz & Nash, 1997). That is, the more experienced an individual is, the more positive attitudes they are likely to develop about computers, which increases self-efficacy levels (Comber et al., 1997). Similarly, Fagan, Neill & Wooldridge (2004) point out that increased computer experience is positively related to computer self-efficacy. In speaking about self-efficacy, Kurbanoglu (2009, p.2) notes that "There is a close link between attitudes and experience, and the attainment of self-efficacy. Research by Bandura (1986) shows that efficacy perceptions develop from a gradual attainment of skills and experience over time".

However, we should note that increased computer experience does not always denote increased computer self-efficacy. That is, a high level of computer usage does not necessarily translate into increased computer knowledge (Sam, Othman & Nordin, 2005). This view is also stressed by Khorrami-Arani (2001, p.17) who points out that "computer experience alone is not the only precursor to student success with computers. Computer knowledge and attitudes also play an important role".

As for the effect of gender on attaining self-efficacy, several studies have concluded that male students have a higher degree of self-efficacy than female students (Betz & Hackett, 1981; Busch, 1995; Hackett, 1985; Isiksal & Askar, 2005; Post-Kammer & Smith, 1985). In Askar & Davenport (2009) study that aimed to explore the factors related to self efficacy for Java programming among first year engineering students. They found that females have less self efficacy than males. Also they states "11.8% of the variance in self efficacy was explained by computer experience" (p. 26). We can conclude then that although computer experience is not the only determinant of computer self-efficacy, it does affect the extent to which self-efficacy is attained and perceived by individuals.

4.3.3. Students and Staff Attitudes Towards CAA

As for students' attitudes towards computer-based tests in particular, Fulcher, (1999) has pointed out that lack of familiarity with computers could have a direct effect on examinees' perception of computerized testing as having little or no experience of using the computer or internet could account for developing negative attitudes towards CBT. Several other studies have investigated students' and academics' attitudes towards computer-based testing, its advantages and disadvantages (Hodson, Saunders & Stubbs, 2002; Sheader, Gouldsborough & Grady, 2006).

Hodson, Saunders & Stubbs (2002) have investigated staff's perceptions of CAA and their views towards introducing it to Glamorgan University. Ten academic staff members (8 males & 2 females) were interviewed by the researchers. The results of the questionnaires reveal that academic staff's main concerns revolve around such pedagogical issues as question types, staff time and staff support.

Sheader, Gouldsborough & Grady (2006) reveal that academic staff's perceptions of CAA are rather positive. "Various advantages of CAA were mentioned by staff members: notably, the reduction in marking time and reduction of paperwork as well as the potential for the software to detect plagiarism and to administer anonymous marking" (p.174). The results of the study conducted by Sheader, Gouldsborough & Grady (2006) also show that the academics who were interviewed cited few disadvantages which they thought would be overcome by training and by improving the existing versions of CAA software. "The use of CAA has proved to be a welcome addition to the tools available to staff members for the assessment of practical classes, and future improved versions of the software will increase the utility of this assessment method" (p.174).

McKenna's study (2001) provides a deeper analysis of academic staff's perceptions of CAA. The participants had different and sometimes mixed attitudes towards the application of CAA. While some perceived it as a substantially important learning and assessment tool, others did not seem motivated enough about its introduction and application in higher education institutes. According to this study, CAA is mainly desired by academics for the following reasons:

- Promoting regular learning behaviour.
- Enabling academics to assess students on a broader scale than is usually possible with the traditional pen-and-paper testing mode.
- Saving time and particularly because of the rapidly expanding number of higher education students.

Other participants, however, have some reservations about the question types used in CAA which they perceive as rather unsuitable for their subjects (participants from humanities and social sciences departments).

Concerning students' attitudes, several comparative studies have been carried out to explore students' perceptions of computer-based testing Students' views were inferred through face-to-face questionnaires or by getting academics to report on their students' views about computerized tests. The results show that *in most cases student perception of CAA is rather positive, and that they would* prefer it to paper-based assessment.

For example, before fully introducing CAA to the Faculty of Life Sciences (University of Manchester), a study was conducted by Sheader, Gouldsborough & Grady (2006) to explore staff and student perceptions of computer-assisted assessment for physiology practical classes. The results of the study reveal that students are willing to implement computer-assisted assessment A similar result was achieved in McKenna's (2001) study in which it is pointed out that the feedback obtained from the academic staff members responding to the National Survey Questionnaire showed that students perceive CAA positively and students would rather take CAA than sit for a traditional paper-based test

In relation to gender differences, there are several studies which concluded that females have more negative feeling towards computerized test than males do (Broos, 2005; Graff, 2003; Houtz & Gupta, 2001; Shashaani & Khalili, 2001). However, several other studies found no gender difference when females have equal exposure to computers (Claraina & Wallace, 2002; Wallace & Claraina, 2005).

Fan ovi ová & Prokop (2008) conducted a study to explore " *students' attitudes towards computer use in Slovakia*". They used a questionnaire consisting of 35 items and divided into three dimensions (cognitive, behavioural and affective) They found that boys have higher mean than girls in the behavioural dimension, whereas no gender differences were found in the cognitive and affective dimensions.

4.4. Evaluation of the Literature

It can be concluded from the literature review that there are inconsistencies in the study's findings. The inconsistency in these findings indicate that mode comparability studies involve different factors that affect the results such as test design, participants, and items content (Pearson,2009). This view is supported by Kingston (2009) who conducted a Meta analysis for studies from January 1997 to March 2007. Some of this inconsistency can be explained as Kingston (2009) suggested by two reasons. Firstly, because of the development of computer administration software that makes the computer mode easier than before for its ease of use design. The second reason is because most of students assigned randomly to take part in either computer or paper mode, so students with more computer familiarity may have decided to choose the computer tests. Moreover, the literature findings can be explained by the differences in the students, differences in the subject or difference in items characteristics (Kingston, 2009).

Many studies concluded that computer familiarity could be considered a major factor which might impact student performance on CAA (Goldberg & Pedulla, 2002). Also there are many factors that affected students performance such as computer self efficacy, attitudes towards computerised assessment, test design and type of questions. So exploring these factors might lead to explain the comparability's results in depth and clarify the reasons standing behind these finding.

Similarly, Wolfe & Manalo (2004) have indicated that low experience with computers would lead to an increased anxiety level, which might give rise to construct-related problems, for students with a lower computer experience (and definitely lower comfort level) would have to face an additional cognitive demand when they are assessed in the computer-based assessment mode. This would create validity-related problems. Hence, the interference of the medium effect would make the assessment less valid However, as Peak (2005) concluded that if CAA used simple MCQs then CAA will be as easy as traditional assessment to the students. Also Wolf & Manalo (2004) results have revealed that none native speakers of English who have a low English language proficiency level would perform better in the paper test than in the computer one, so anticipating that Omani student, especially those who had low level of English language, will be disadvantages from CAA

In spite of the importance of the demographic factor such us gender, region, ethnic minorities, in these comparability studies, there are just a few studies which have been conducted in this interesting area (Gallagher, Bridgeman & Cahalan, 2002; Parshall & Kromrey, 1993). In most cases, studies have yielded small differences between subgroups and in few cases; no significant themes were actually traced. Horkay et al., (2006), for example, have revealed no significant differences between delivery modes and no significant interaction between direct writing assessment and any demographic factors like gender, race, (or ethnic background), parents' educational level, school location and school type, despite what has been noted by some studies that women, Africans and Spanish speakers are less likely to have access to computers (Miller & Varma, 1994; Reinen & Plomp, 1993). Moreover, Wolfe & Manalo (2004) reveal that many studies have indicated that both females and ethnic minorities have less access to computers and this would diminish their proficiency level in computer-related tasks. Therefore, Wolfe & Manalo (2004) conclude that both females and ethnic minorities tend to experience higher anxiety and lower confidence levels in all the tasks that are related to the computer. Hence, according to Wolfe & Manalo, female students and the students coming from developing countries tend to choose the paper-based administration mode rather than the computer-based one, when they have the possibility to choose. However, the question is, as all these studies were not conducted in these developing countries, is comparability studies will yield different result if conducted in the developing countries?.

The gender differences in CAA has received a greet attention in the comparability studies. Many of these studies have shown that there is no gender difference in students' attitudes towards computer tests (Baggott & Rayne, 2001; Beverly et al., 2001; Busch, 1995; Choi & Tinkler, 2002; Horkay *et al.*, 2006; Horne, 2007; Kies, Williams & Freund, 2006; McKenna, 2001). Moreover, students' attitudes towards CAA tend to be rather positive and most students even prefer this mode of assessment to the traditional paper-and-pencil one. However, Omani students have different characteristics as they come from different regions and have different computer experience. So this research study may arise different results compared the studies in the literature.

4.5. Conclusion

There are inconsistent findings in the literature which aimed to compare between two modes of administration (paper versus computer mode). Hence, it makes it difficult to generate a clear conclusion which based on previous studies. Most recent studies concluded equivalence between the two modes

(Texas Education Agency, 2008; Wang et al., 2007, Wang et al., 2008). Nevertheless some of these studies do not focus on the different participant variables and dealt with them as one. The conclusion we can draw from investigating the literature review on mode of administration comparability studies can be summarized as follows:

1. There are very few studies which have been conducted to explore the impact of testing modes on students from a demographic perspective by focusing on such variables as ethnicity, geographical region and social status. However, the studies which have addressed this subject conclude that there is hardly any difference in performance scores among subgroups. In Omani context , as we mentioned earlier, Oman consists of nine region which is different to some extant in the availability of computers, also Omanis have various regional specificities so, it is difficult to decide on the comparability of both assessment modes by solely relying on the findings of the studies addressed in this chapter.

2. Most studies have focused on students attitudes while few studies have investigated academic staff's attitudes towards computerized test administration.

3. Attitudes towards CAA are mostly positive in most of the studies presented in this chapter. Moreover, some studies concluded that most students even prefer computerized assessment to the traditional paper-and-pencil one. However, is short computer history use will affect Omani students' attitudes towards CAA?

4. Most of the studies concluded that males have more computer self efficacy than females even in case that they have similar computer experience.

5. Many studies concluded that females have more negative feeling towards tasks related computers compared to males.

6. Very few studies are the ones which have focused on Middle East student cohorts, or addressed students' English language proficiency levels, especially in placement tests.

7. Many of the studies presented have indicated that those who have more experience with computers perform better in CBTs than those who have less experience with computer use. However, several studies have concluded that the computer experience factor has a negligible effect on the students taking multiple-choice questions.

So this current research study seeks to contribute to the use of CAA field. That is, our intention is to assess whether Omani students performance scores would ever be affected by such a mode of delivery change and if so what the factors affecting students performance.

Chapter Five: Study Purpose, Method and the Initial Validation

5.1. Introduction

As assessment is an effective variable in the educational process which plays an important role in both teaching and learning, then changing it needs to be based on various factors. That is before changing any assessment method, it is necessary to identify the factors that might influence its implantation and to ensure both the quality and effectiveness of the new method. However, this change cannot be fully achieved unless a new assessment culture that is based on trust among the different educational parts is built, and this could only be achieved through identifying the potentials of the new assessment method and the advantages it has over the traditional assessment one.

This study will compare students' performance across both ways of assessment (paper-pencil assessment and computerised assessment) in terms of such variables as gender, colleges and region of residence. Moreover, it will investigate the effect of computer experience and computer self efficacy in the light of the same variables focusing on the relationship between students' performance and their computer experience and computer self efficacy levels It will also try to investigate the students' attitudes, views, thoughts and perception towards computerised assessment Similarly, this study will investigate whether or not computer experience and computer self efficacy have got any effect on students' attitudes towards computerised assessment, and also explore the relationship between students' attitude and their performance. In addition, it will explore the academic staff thoughts and feelings in depth towards CAA as well as the difficulties and limitations which might be faced when implementing CAA from their points of view and the suggestions to overcome these difficulties.

Also it aimed to explore the advantages and disadvantages for CAA from the staff point of view.

5.2. Study Questions and Hypotheses

This study will try to answer the following questions:

1- Do students perform differently on computer based testing versus paper based testing?

2- Does test mode differentially affect the performance of student groups those categorized by gender, colleges, or by regions of residence?

3- Do computer experience and computer self efficacy vary between variables such as gender, colleges and region of residence?

4- Is there a relationship between students performance, and computer experience and computer self efficacy?

5- What are the students' perceptions, thoughts, views and feeling towards implementing computerised assessment?

6- What is the effect of the different variables such as gender, colleges and regions on the students' attitude towards computerised assessment?

7- Is there a relationship between students attitude towards computerized test, student performance, computer experience, and computer self efficacy?

8- What are the academic staff's perceptions, thoughts, views and feeling towards implementing CAA?

9- What is the effect of gender, nationality and the mother tongue language on the academic staff's attitude towards CAA?

10-What are the difficulties and limitations that might face CAA use in Oman from the academic staff view?

11-What the CAA advantages and disadvantages from the academic staff point of view in Omani higher education institutions?

The general hypothesis of this research is that, when comparing students' performance in both paper-pencil test and computerized one, will the results be different with different mode of assessment Based on the general hypothesis and the nature of the planned study; the study hypotheses are as follows

Hypothesis 1: There is no difference between average test scores in relation to the method of delivery.

Hypothesis 2: There is no difference between average test scores in relation to examinee's gender, college of study and region of residence.

Hypothesis3: There is no difference on computer experience and computer self efficacy in relation to examinee's gender, college and region of origin.

Hypothesis4: There is no relationship between computer experience, computer self efficacy and student performance.

Hypothesis 5: There are no differences between students' attitudes towards Computer Assisted Assessment in relation to the gender, college, region of origin.

Hypothesis6: There is no relationship between student attitudes towards computerised assessment, and students' performance.

Hypothesis 7: There are no difference between academic staff attitudes towards CAA in relation to their gender, nationality and mother tongue language.

In order to improve the validity of the study, a validation phase was conducted as the first phase. This phase aimed to validate all study instruments and determine if any changes are necessary. Thus, this chapter describes the approach which this research followed which includes the instruments which were used and as well as their description. Then the result of the validation phase study will be described. Finally, the research design and procedures for the main study will be given.

5.3. Research Method of Data Collecting, Variables and Instruments Description

5.3.1. Research Method of Data Collecting

This study uses multiple methods (quantitative and qualitative) to get a full picture of the effect of mode administration and the factors that might affect students' performance. The main advantage of employing multiple methods is, as commonly cited, to permit "triangulation" (Robson, 2002, p.372). That is, the use of multiple methods is meant to enhance interpretability through data triangulation. For example, in a primarily quantitative study, the interpretation of the statistical analysis may be enhanced by a qualitative narrative account. Conversely, "a qualitative account may be the major outcome of a study but it can be enhanced by supportive quantitative evidence used to buttress and perhaps clarify the account" (Robson, 2002, p.372). This combination of methods will allow the triangulation of the study findings and will either confirm the conclusion or identify inconsistencies. Also data triangulation will be used to try to enhance study findings and make them more comprehensible and more consistent. In addition, triangulation can help to counter some of the threats to validity and so increase the trustworthiness of findings.

The instruments that will be used to collect data quantitatively are the test and questionnaire. The instruments that will be used to collect data qualitatively are focus groups and interviews. The diagram below (figure 5) illustrates the method and instruments used to collect study data.

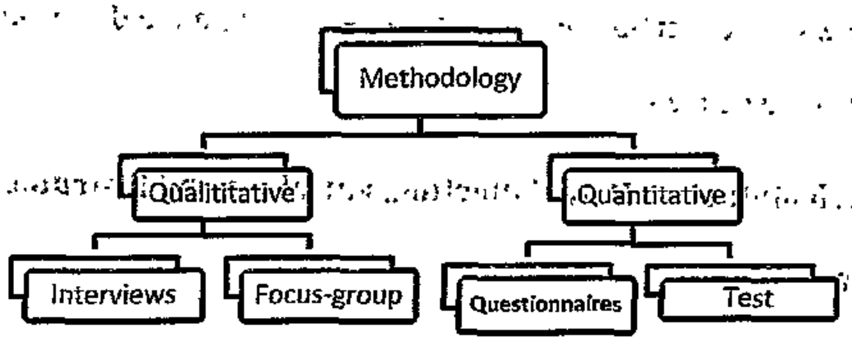


Figure 5. Method and Instruments Used in This Research.

5.3.2. Study Variables

The main outcome measure is the student test score. There are many auxiliary variables that will help us to understand variation in test scores. For students these include, gender, college, region.

For staff, the background information will include gender, nationality and the mother tongue language. For more clarification the different variables are listed below:

Student variables

- The Outcome measure is : Test score

- The experimental variable is: Administration mode (paper test or computerized test)

The other explanatory variables fall into two groups: demographic or social factors; and psychological factors.

- The demographic or social variables are:
  1- Gender (male or female)
  2- College (place of study)
  3- Region (place of residence),
- The psychological factors are:
  1- Computer experience
  2- Computer self efficacy
  3- Attitudes towards Computer Assessment Scale score (ATCAS).

Staff variables

- The demographic variables are
  1. Nationality and language (Omani, English Speaking non-Omani or Arabic Speaking non-Omani),
  2. Gender (male or female).
- The psychological factors are
  1. Attitudes towards Computer Assisted Assessment (CAA) score.

5.3.3. Study Instruments

The following instruments were selected to collect data relevant to the research questions

Student instruments

- English language placement test.

- Demographic data questionnaire.
- Computer experience questionnaire.
- Computer self - efficacy questionnaire.
- Attitude Towards Computerised Assessment Scale (ATCAS).
- Pre- test and Post -test Focus Groups.

Staff instruments

- Attitudes towards Computer Assisted Assessment (CAA) questionnaire
- Semi-Structured Interviews.

Now each instrument used to collect data in the study will be described, were the entire instruments are placed in Appendix A.

5.3.4. English Language Placement Test

This study will use an English language test because most Omani higher education institutions have a foundation year, which requires students to attain a certain level in English language proficiency (which is equal to band 4.5 in the IELTS test) before being allowed to start studying the subject they choose. So an English language test is a suitable one which will represent the students' performance in all Omani higher education institutions.

The English language test was prepared by the Assistant Director of the English programme in the Ministry of Higher Education. This department is responsible for preparing, scoring and analysing the final English language tests for the all Omani Applied Science Colleges foundation year. The English programme director

- *Takes responsibility of the effective functioning of the degree and foundation* programs.

- Liaises with the heads of English departments to ensure consistent test administration procedures, standardized assessment and reporting.
- Liaises with the other program directors to ensure the English program fulfilment of the students' language needs in the degree programs.
- Liaises with other academic institutions in Oman and internationally for information sharing and development and evolution of the English program.

Test Content

The English language test consists of two parts; the first part is developed to test grammar and comprises 30 multiple-choice items, each containing four alternatives. The second part is devoted to reading and consists of three short passages with 4 items for the first passage and 8 items for the second and third passages (20 items total). The question types in the reading part consist of MCQs, gap-filling, matching and true/false questions. Before the test was administered on computer, the paper-pencil version was converted to a computerized format.

Paper-Pencil Test

In the paper-pencil test the items were presented jointly on a page. In the reading part, all the passages were presented on one page followed by test items. The test-takers responded to the items on a separate answer sheet. In addition, they were able to move freely throughout the whole English test (grammar and reading). They could respond to items in any order but they had to give a response to all items.

Computerized Test

In the computer presentation, the grammar test items were presented jointly on the screen. There were 20 grammar items shown in a single screen at the same time. Reading test items appeared immediately after grammar items. In the reading test, each of the three passages appeared with its items on a single screen After finishing the test students were able to review their answers and change some answers if they wanted So, there were two options shown on the screen at the end of the exam. One is "*save and review*" and the other is "*finish and submit*".

5.3.5. Demographic Data Questionnaire

Because this study requires knowing the demographic and social backgrounds of the participants; students were asked to fill in a demographic questionnaire reporting their academic number, gender, college, region of residence and whether they have computer at home or not (See Appendix A). The reason for this is to match students' demographic data with their test results. Students who were sitting the computerized test were also given a unique user name and password. Linkages between responses to questionnaire, test result as well as demographic and social background were established using either the unique username or the academic number

5.3.6. Computer Experience Questionnaire

Computer experience has been shown to be an important variable affecting performance in CAA studies. Many studies concluded that lack of computer experience could cause equity and validity issues. That is, students' achievement might be affected by other variables than the test's content.

However, these research strategies had different ways of measuring computer experience, and many of them have actually focused on measuring computer experience just from one perspective, and that simply the frequency of computer use. However, there are several studies which have linked computer experience and perceived knowledge of computers to computerized test performance (Smith, 2003). Smith (2003) discussed the difficulty of inferring the effect of computer experience on performance in the computerized test and the equality of the tests, because most of the studies did not provide either a specific and accurate definition of computer experience or a potential way to measure it. In a study by Smith et al., (1999) computer experience has been defined as frequency of use, time of use, kind of use, numbers of computer courses, owning a computer and attitude towards computers. They also highlighted the importance of distinguishing between quantity and quality issues of experience with computers. They have also specified the notion of computer experience reflecting two perspectives both objective and subjective constituents.

Therefore, in this study focus was laid on two aspects of computer experience through the amount or frequency of computer use and the kind of use. Based on the literature review 8 items from the questionnaire on computer experience that had established validity and reliability in other studies are modified and developed to fit into this study (Al-Kother, 1999; Fagan, Neill & Wooldridge, 2004; Johnson, Ferguson & Lester, 1999; Smith, 2003). The first four questions of the questionnaire were meant to reveal the amount of previous use of computers. The remaining four items elicited the kind of use (See Appendix A). All 8 items are rated on a 3 point Likert –type scale which expresses little, moderate and lots of computer experience.

5.3.7. Computer Self-Efficacy Questionnaire

Self efficacy is defined as "the belief in one's ability to execute successfully a certain course of behaviour "(Busch, 1995, p.147). Whereas 'computer self-efficacy' has been defined as "an individual belief in their ability to use technology *in order to solve problem, make decisions, and to gather and* disseminate information" (Stephens & Shotick, 2002, p.591).

Most research studies support the idea that self-efficacy is an important variable to predict an individual's behaviour (Busch, 1995). In addition, Torkzadeh, Koufteros & Pflughoeft (2003, p.263) point out "computer self efficacy plays a key role in self motivation and affects potential usage of information system technology". They also indicated that few studies were conducted on the effect of computer self efficacy. According to Torkzadeh, Koufteros and Pflughoeft (2003), " in general, researchers to date confirm that computer self-efficacy not only determines decisions by individuals to accept and use the computer system, but is also a good predictor of achievement in computer-related tasks" (p.264). Also Bandura (1977) asserts that self efficacy expectation is an important determinant of anyone' success in overcoming any hurdle, for it determines how many attempts will be made and how much effort will be demonstrated to face the task difficulties

So, in order to explore student computer self efficacy, seven items from the computer self efficacy questionnaire which was originally developed by Fagan, Neill & Wooldridge (2004) were used The internal consistency they reported for the scale was 0 93. All 7 items are rated on a 5 point Likert –type scale (See Appendix A). The response was given according to the following criteria: 1= very little confidence, 2=little confidence, 3= some confidence, 4= moderate confidence, 5= high confidence

5.3.8. Students Attitude Towards Computerised Assessment Scale (ATCAS)

Many studies have investigated students' attitude towards computers in general (Fan ovi ová & Prokop, 2008). However, relatively few have investigated student attitudes and perceptions towards CAA (Baggott & Rayne, 2001; McKenna, 2001; Sheader, Gouldsborough & Grady, 2006; Smith & Caputi, 2004). Therefore this study aimed to explore students' attitudes towards computerized assessment as this factor is very important to explore before shifting to CAA in any institution. So, in order to measure the attitudes toward computerised assessment, a Likert type attitude scale which was validated and used in a number of prior published studies was used (Smith & Caputi, 2004). Specifically, "the attitude towards computer assessment scale (ATCAS) was developed to allow for the exploration of examinees reactions towards computerized relative to conventional testing methods" (Smith & Caputi, 2004, p.409). ATCAS consists of 13 items. Nine of the items "were taken from an attitude questionnaire developed by Burke et al *(1987)*" (cited in Smith, 2003, p.106). All 13 items are rated on a 5 point Likert-type scale ranging from 'strongly disagree' (1) to 'strongly agree' (5). Higher scores reflect a greater preference for conventional testing method. However in this research I reversed coded to let higher scores reflect higher preference for the computerized test. The internal consistency reported by Smith (2003) was good (alpha of 0.79).

5.3.9. Pre – Test and Post -Test Focus Groups

To expand upon our findings from the *ATCAS* questionnaire, focus group discussions were also conducted. Focus groups are a very effective method to use to explore views, experiences, feelings and thoughts on specific issues. Morgan (1988, 25) points out that "focus groups are useful when it comes to investigating what participants think, but they excel at uncovering why participants think as they do". Powell & Single (1996) defined a focus group as "a group of individuals selected and assembled by researchers to discuss and comment on, from personal experience, the topic that is the subject of research" (p.499)

So, students' focus groups were conducted before and after getting them to sit for the computerised test Participants were requested to explain the reasons behind their views so that the researchers could identify and record such interactions. Moreover pre-test focus group meetings will give us a view of students' feelings and perceptions of CAA, especially if they do not have any experience in it. However, any transformation in their feelings (if it happens) will be noticed during the post- test focus group meeting.

The aim behind doing this was to get a real insight into students' feelings and thoughts as well as the potential factors affecting their perceptions. As interaction between participants in focus groups is important, a small-size focus group is a vital element to enable and encourage participants to interact with each other, which uncovers their ideas and perception. The recommended number of individuals in a focus group is usually six to ten (O'Connell & Dyment, 2006). Therefore, a focus group was selected for this study in order to draw on the students' attitudes, feelings, beliefs, experience and reactions towards assessment by computer in a way which would not be possible using solely other research methods, such as questionnaires or one to one interviewing (Krueger, 1988; Morgan, 1997; cited in O'Connell & Dyment, 2006).

A list of 11 questions was used to guide the interview (See Appendix A) The prompt questions were developed and modified based on the literature review (Sheader, Gouldsborough & Grady, 2006). That is, the questions were designed in a way which allows the researcher to gain more insight, from students' perspective, into the themes emerging from other instrument finding of the study. The session was recorded on audiotapes and later transcribed.

5.3.10. Staff Attitudes Towards Computer Assisted Assessment Questionnaire (CAA)

In spite the substantial number of studies carried out to investigate students' attitudes towards computers and Computer Assisted Assessment, very few studies have been conducted to investigate staff attitudes towards CAA. The Sheader, Gouldsborough & Grady (2006) study was one of the few studies that explored academic staff member perceptions and views towards CAA using a questionnaire and face to face interviews.

So, in order to measure Omani staff attitudes toward Computer Assisted Assessment (CAA) and based on the literature review that investigated the staff and student attitude towards CAA (Baggott & Rayne, 2001; McKenna, 2001; Sheader, Gouldsborough & Grady, 2006; Smith & Caputi, 2004) Likert type attitude scales was developed from the literature and validated as part of this study (Appendix A). The questionnaire consists of 18 items which are rated on a 5 point Likert type scale ranging from 'strongly disagree' (1) to 'strongly agree' (5). Six items (from 9 to 14) were reverse coded, so that higher scores reflect more positive attitudes towards CAA.

5.3.11. Staff Semi-Structured Interviews

Face to face semi-structured interviews were scheduled and conducted with Omani and non-Omani academic staff, both males and females. The semi-structured interviews were meant to render the interview flexible and create a stress-free environment for the interviewees, which enables the interviewer to explore in depth any details or topic-related issues without being tied to a rigid interview schedule (Davis-Case, 1990).

A list of 18 questions was developed and modified based on the literature review (McKenna, 2001; Sheader, Gouldsborough & Grady, 2006) and used to guide the interviewer to find answers to the questions along with the staff questionnaire (See Appendix A). The questions were divided into three parts. Before starting, the interviewed academic staff members were given some background information about Computer Assisted Assessment (CAA) and about the main aims of the interviewer. The first part of the interview included questions about their computer experience in general and how often they use a computer at work. It also included questions about their background information about computerised tests. The second part included questions to explore the advantages and the disadvantages of computerised tests from their point(s) of view, as well as the difficulties and limitations that might be faced by any higher education institution intending to apply CAA, and their suggestions to overcome such hurdles. The third part aimed to explore their potential reaction(s) if their institute intended to switch to computerised test, and the most important worries they would have about that switch. This part aimed to investigate their willingness or possibly their readiness to attend any training sessions in CAA offered by their institutions if computerised testing was to be adopted by that institution.

5.4. Validation Study

Although the adopted research instruments were based on current evidence, it was important to validate them through field testing in circumstances specific to this study This section describes the phase one study which aimed to validate all study instruments before conducting the main study. There are two main reasons for the validation study. First, it is intended to validate the instruments and to modify them if necessary. Second, it seeks to investigate the feasibility of conducting the research and explore the likelihood and nature of any organisational problem which could be encountered in the main study.

The validation study was run from January to March 2008 based on one Applied Science College (Nizwa College). The reasons behind choosing this particular college are firstly, because it had students from different regions. Secondly, it was geographically the closest college to the researcher, which made direct communication with the administration easier, particularly in following up with the necessary arrangements for this phase. Also, it is easier to reach this institution and get any missing information or data in case it is needed. [Most importantly, however, was the fact that the findings of the first phase (validation study) were used to refine the research questions and the used method for the main research study].

The evidence of the validity of the instruments was accumulated from many sources. Evidence based on internal structure was accumulated through the use of the principal component analysis of all questionnaires. Also, evidence based on test content was accumulated by giving the instruments (test, questionnaires, focus groups and Interview questions) to a group of experts to assess their suitability and judge whether the instruments are likely to measure what they were intended to measure. All instruments (test and questionnaires) were given to a number of academic staff working in the Applied Colleges and having good knowledge about CAA. Most feedback received concerned spelling mistakes, rather than necessary changes to the items.

The use of the terms validity and reliability are common in quantitative research. In terms of qualitative research the validity concept is replaced by "*credibility*" whereas reliability is replaced by "*dependability*" (Guba & Lincoln, 1981). Both credibility and dependability lead to increased research "trustworthiness". According to Seale (1999, p 266) "trustworthiness of a research report lies at *the heart of issues conventionally discussed as validity and reliability*"

In addition, the "*triangulation*" was used as source of validity evidence or credibility in qualitative data as well. Creswell & Miller (2000, p.126) cited in Golafshani (2003, p.604) defined triangulation as "a validity procedure where researchers search for convergence among multiple and different sources of information to form themes or categories in a study".

5.4.1 Student Sample in the Validation *Study*

116 foundation year students attending Nizwa College participated in the validation study. Classes were chosen randomly and students in each class were asked if they were willing to participate in the study. 68 (59%) students sat for the paper test while 48 (41%) students the computer test. 65 (56%) students were males and 51 (44%) were females. Table 3 below shows the *number of the students according to gender, mode and region.*

Mode	Gender	Region				Total Gender group
		SHARQIYA	BATINAH	DHAHIRA	DAKHLIA	
Paper Mode	Male	9	12	7	12	40
	Female	6	4	7	11	28
	Total paper group	**15**	**16**	**14**	**23**	**68**
Computer Mode	Male	5	9	4	7	25
	Female	4	3	2	14	23
	Total computer group	**9**	**12**	**6**	**21**	**48**

Table 3. Students Sample in the Validation Study According to the Mode, Gender and Region of Residence

For the focus groups, 10 volunteer students (6 male and 4 female) volunteered to participate in a pre–test and post- test focus group interview. Only students who sat the computerized test participated in the focus group.

5.4.3. Staff Sample in the Validation Study

52 randomly chosen academics participated by filling in the questionnaire. 18 of them were females and 34 were males. The number of Omanis was 8 while the number of non-Omanis was 44. In addition, four academic staff (2 male and 2 female) were interviewed. Table 4 below shows the staff number according to their gender and nationality.

Language	Gender	Nationality		Total Staff Sample by Nationality/ Language	The Total Staff Sample
		Omani	Non Omani		
English	Male	-	17	28	52
	Female	-	11		
Arabic	Male	5	12	24	
	Female	3	4		
Total staff sample by Gender	Male	5	29	34	
	Female	3	15	18	

Table 4. Staff Sample in the Validation Study According to the Gender, Nationality/ Language

5.4.3. English language Test Validity and Reliability

Evidence based on test content was the source of the validity evidence of the English language test depending on the judgement of experts from the English departments at the Applied Science Colleges. That is, before being administered, the test was given to a number of academics from English departments for assessment. Also, the English language test itself (see

Appendix A) was prepared by an academic staff member working as assistant director in the English language programme departments.

In relation to reliability, Cronbach's Alpha was used to explore to the entire sample as well as the paper and computer groups in a separate way. The English Language test reliability was 0.86 for the paper version and 0 78 for the computer one. The reliability of the grammar and the reading tests was good in both modes. These coefficients indicate satisfactory internal consistencies for the test. Table 5 illustrates the reliability coefficient for the total test as well as the grammar and the reading tests.

Instruments	**Reliability coefficient**	
	Paper test	***Computer test***
Total test	*0.86*	*0.78*
Grammar	*0.81*	*0.74*
Reading	*0 66*	*0.54*

Table 5 Reliability Coefficient for the Entire Test, Grammar and Reading

5.4.4. Validation of Computer Experience Questionnaire

All 116 foundation year students completed the questionnaire. The eight items were subject to principal component analysis. The Kaiser-Meyer-Oklin (KMO) value was 0.74 exceeding the recommended value of 0.60 (Kaiser, 1974) and the Barlett's test of sphericity reached statistical significance supporting the factorability of the correlation matrix.

Factor analysis revealed that most items (5 out 8) loaded in one factor which reflected the amount of computer experience and explained 32.96 percent of the variance. Two other items (3, 7) loaded in the second factor and explained 14.30 percent Just one item (item number eight) loaded on the third factor and explained 13.67. However, inspection to the scree plots below revealed a clear break after the first component (Figure 6).

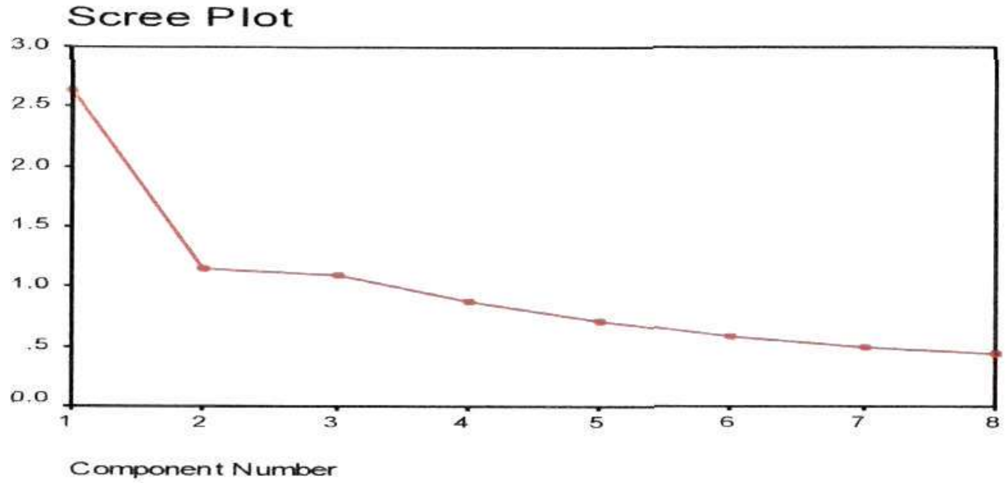


Figure 6. Scree Plot of Computer Experience Questionnaire in the Validation Study

It was decided to retain one component for further investigation. The one component solution explained a total of 32.96 percent of the variance. For most items loading on this component the correlation was more than 0.30 except one item (item number 7) which had a low correlation (0.15). The internal consistency of the scale (Cronbach's alpha=0.66) was acceptable. In addition, after deleting item number 7, the reliability coefficient was increased to 0.69. The item-item correlations ranged from 0.22 to 0.58 except for item 7 for which the correlation was very low (0.08). Moreover, the squared multiple correlations (SMC) ranged from 0.11 to 0.39, which provided evidence for the internal consistency of the computer experience scale.

As the content of item number seven in the computer experience questionnaire is related to playing computer games ("Do you use computers to play games?"), its main aim is actually to explore student prior computer experience and use. However, it has been decided to leave this item pending until the number of the *participants (student sample) increases to include the whole study group* members in the main study, which might change the overall results.

5.4.5. Validation of Computer Self-Efficacy Questionnaire

All 116 foundation year students completed the questionnaire. The seven items were subject to principal component analysis. The Kaiser-Meyer-Oklin value was 0.93 exceeding the recommended value of 0.60 (Kaiser, 1974) and the *Barlett's test of sphericity reached statistical significance supporting the* factorability of the correlation matrix.

The analysis revealed that all seven items loaded in one component with high correlation ranged from 0.82 to 0.87. The total variance explained was 72%. Moreover the scree plot reveals a clear break after the first component (Figure 7).

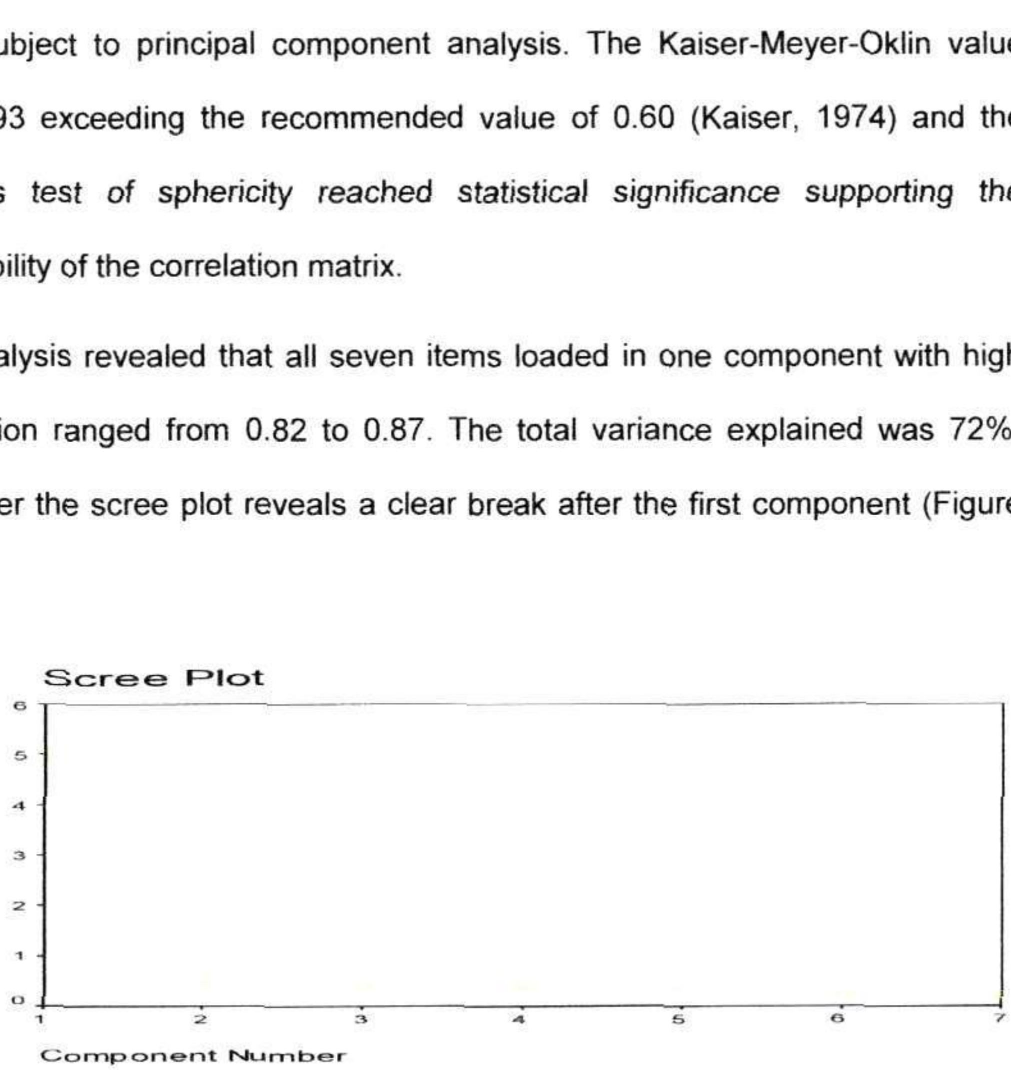


Figure 7. Scree Plot of Computer Self-Efficacy Questionnaire in the Validation Study

The internal consistency of the scale (Cronbach's alpha=0.93) was good and similar to that found by Fagan, Neill & Wooldridge (2004). Moreover item-item correlations were in the expected range of 0.60 to 0.70, indicating homogeneity among scale items. The squared multiple correlations (SMC) ranged from 0.58

to 0.69. Together these findings provide further evidence for the internal consistency of the computer self efficacy questionnaire.

5.4.6. Validation of Students Attitude Towards Computerized Assessment Scale (ATCAS)

Only the 48 students who had sat the computerized test participated in completing this questionnaire. The sample consisted of 25 males and 23 female students.

Smith & Captui (2004) analyse the questionnaire by factor analysis using a sample of 82 students. Their result revealed two separate factors (Eigenvalues greater than one) .The first factor consisted of 9 items and was labelled ease of use. The second factor was labelled CBT confidence and consisted of 4 items.

In the present study, the Kaiser-Meyer-Oklin value was 0.60 meeting the recommended value 0.60 (Kaiser, 1974) and the Barlett's test of sphericity reached statistical significance supporting the factorability of the correlation matrix.

Principal components analysis revealed the presence of five components with eigenvalues exceeding 1 explaining 23.83 percent, 13.82 percent, 11.02 percent, 9.66 percent and 8.53 percent of the variance respectively. The total variance explained was 66.88 percent.

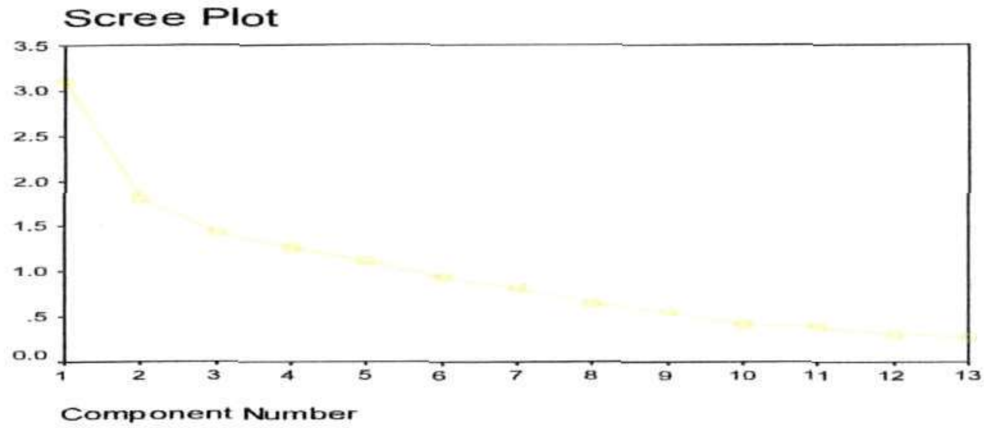


Figure 8. Scree Plot of the ATCAS in the Validation Study

However, an inspection of the Scree plot revealed a break after the first component (Figure 8). So, it was decided to retain one component for further investigation. The one component solution explained a total of 23.83 percent of the variance. For most items loading on this component the correlation was more than 0.30 except four items (4, 5, 7 and 10) which had low correlation, 0.29 for items 4, 5 and 10 and 0.22 for item number 7. Moreover item -item *correlation ranged from 0.13 to 0.58 and the squared multiple correlations* (SMC) ranged from 0.22 to 0.48. The internal consistency for the scale was 0.70 which considered acceptable. Together, these finding provide reasonable evidence of the internal consistency of the ATCAS.

5.4.7. Validation of Staff Attitude Towards Computer Assisted Assessment

52 academics staff members from Nizwa College participated in completing the questionnaire. The sample consisted of 18 females and 34 males. The Kaiser-Meyer-Oklin value was 0.69 exceeding the recommended value 0.60 (Kaiser, 1974) and the Barlett's test of sphericity reached statistical significance supporting the factorability of the correlation matrix.

Principal components analysis revealed the presence of five components with eigenvalues exceeding 1 explaining 29.05 percent, 12.72 percent, 11.35

percent, 7.63 percent and 6.842 of the variance respectively. The total variance explained was 67.61% percent.

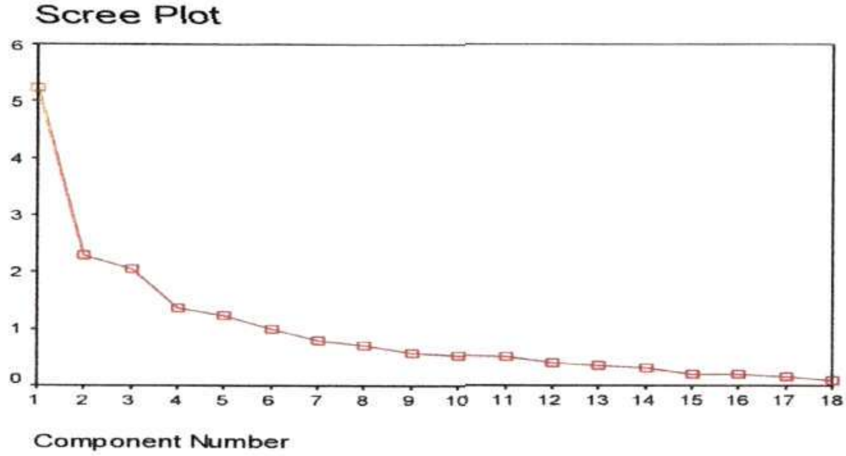


Figure 9. Scree Plot of the Staff Questionnaire in the Validation Study

However, an inspection of the scree plot revealed a clear break after the first component (Figure 9). It was decided to retain one component for further investigation. The one component solution explained a total of 29.05 percent of the variance. For most items loading on this component the correlation was more than 0.40 except four items had low correlation [item number 7(0.15), 9(0.29), 13(0.12) and 11(0.39)].

The internal consistency of the scale was acceptable (Cronbach's alpha 0.57). In addition the item-item correlation was 0.11 to 0.45 except item number 14 which had low correlation (0.02) and the squared multiple correlations (SMC) ranged from 0.31 to 0.79. Together, these findings provide some evidence for the internal consistency of the scale.

5.4.8. Validation of the Students Focus Group Method

The focus group method was tested through the phase one study using one college with 10 students who volunteered to participate in the pre-test and post-test focus group interviews. Focus group methods have been chosen because they provide a flexible method for interaction between the students and the researcher and among the students themselves, which helps to get a deeper insight into their beliefs, attitudes and values in relation to the research subject. In addition, the questionnaire interview was used to set a deeper understanding of the data. It also helps to explain particular findings from the other instruments used, like the questionnaire, which highlights some issues and trigger further research and exploration. Volunteers for the focus groups have been chosen randomly from the sample of the classes intended to take the computerized test because they are able to provide relevant and accurate information to the researcher before and after sitting for the computerized test. Those students would be the first students to face a computerized test if the institution decided to adopt CAA.

After conducting the focus group method, the audiotapes were transcribed and the data was analysed. The data obtained from the focus group discussion met the study aims, revealing valuable answers, which shows the usefulness of such a tool. Beside the interview, the questionnaire was very useful as it contained most of the questions that were needed by the researcher to direct the focus group discussion. Moreover the triangulation of data which was used increased the credibility of the research data and enhanced the trustworthiness and the quality of the research.

5.4.9. Validation of the Staff Semi-Structured Interview

Most of the academics who were interviewed in the validation study were from the English language department because they are the ones who teach foundation year students, which means that they are more capable than other academics of providing the most accurate information. It is worth noting that the researcher is one of those academics who normally teach in applied colleges and to avoid any sort of bias, I decided not to interfere and rely on the information obtained from colleagues. In addition, the initial results were discussed with two colleagues in order to get further explanations, interpretation and suggestions on data collection and how to deal with it. After listening to the tapes several times, all the data was transcribed and analysed through thematic content analysis. The interview questionnaire was very useful and met most expected research aims. Moreover, method triangulation increased the credibility and the dependability of the research.

5.4.10. Operational Issues

There were two particular operational issues which arose during this validation study, and which were not expected. First, I was advised by the head of the Learning Resource Centre where I held the online tests to get permission from the college dean to switch off the internet from all the college premises except for the LRC. This was to ensure a successful and a smooth performance which would allow me to apply the computerized test without having to worry about internet problems.

Second, with the pre-test focus group, for more than 20 minutes I struggled to convince the female students to join in along with the male ones. At first they refused to sit facing males and some even withdrew and refused to participate, but eventually, I managed to convince them. Besides, during the focus group session, I struggled a lot to get their opinions due to their shyness. However, after completing the computerized test the female students participated in the post-test focus group discussions actively and I did not have to go through what I had gone through with the pre-test focus group so the discussion went on very smoothly. Therefore, in the main study it was made clear from the beginning for female students that those who were ready to volunteer and take part in the focus group would later have to sit in a round table facing males.

5.5. Research Design

Recently, the issue of shifting from conventional to computerised testing has started to acquire much attention, especially in higher education institutions. In the Sultanate of Oman this issue was launched in all universities and colleges, through producing CAA programmes such as Moodle, Blackboard and WebCT. So the main purpose of this study is to explore the possibility of implementing Computer Assisted Assessment in Omani higher education institutions.

After the validation study phase I found that students from different regions might have differences in their computer experience and computer self efficacy levels. So, I made sure that these factors are included in the research design and also tried to get a large sample that actually represented students from the different regions studying at different colleges.

As *Applied Science Colleges are definitely higher education institutions which* are located in different areas and which have students from different regions of the Sultanate, then these colleges provide a good sample of Omani higher education institutions.

This study has two samples. One of them is students and the other is academic staff. For students, an experimental group design was used to examine mode of administration effects. The participants were assigned to participate in this study according to the class allocation. Classes were randomly allocated to the experimental group who sat the computer test and the control group who sat the paper-and- pencil test. The time available for testing was two hours for both groups. For the computer test "Moodle" software was used.

5.5.1. Students Sample in the Main Study

The population of this study was defined as "full time foundation year undergraduate students at Applied Science Colleges in the Sultanate of Oman". There were 1504 students in the five Applied Colleges attending a one-year foundation course. These colleges are located in different regions affiliated to the Omani Ministry Of Higher Education. As College number six does not have foundation year students, table 6 below shows the number of foundation year students in the five colleges only.

College/Gender	Male	Female	Total Population by College
Ibri	73	183	256
Sur	214	97	311
Sohar	319	123	442
Nizwa	106	137	243
Salalah	196	56	251
Total Population by Gender	908	596	1504

Table 6. The Population of Foundation Year Students Regarding Gender and College for 2008/2009 Academic Year

I chose three colleges (Sur, Ibri, and Nizwa) are located in different regions. The main reason for choosing these colleges is that the majority of the students in these colleges are from different regions except one college (Sur) which has students just from one region only (AlSharqiya). However, as these colleges are located in different regions and as it is so far geographically between each other, the students may have different characteristics. The total number of students who participated in the study was 439, of which 212 were females and 227 were males. Table number 7 below shows the number of students according to their gender, colleges and regions

Regarding the focus group sample, there were 10 to 12 students both males and females who participated voluntarily in focus group discussion from each of the three sampled colleges.

College	Gender	AlSharqiya	AlDhahira	AlBatinah	AlDakhilya	Total sample by College	
Sur	Male	84	-	-	-	84	**133**
	Female	49	-	-	-	49	
Ibri	Male	5	45	9	5	64	**163**
	Female	5	44	46	4	99	
Nizwa	Male	18	18	23	20	79	**143**
	Female	14	10	10	30	64	
Total sample by Gender	Male	107	63	32	25	**227**	**439**
	Female	68	54	56	34	**212**	
	Total sample by Region	**175**	**117**	**88**	**95**		

Table 7 The Sample of Males and Females at Different Colleges from Different Regions.

5.5.2. Staff Sample in the Main Study

All academic staff members in the three sampled colleges were invited to complete the questionnaire. *Table number 8 demonstrates the number of* academic staff in all six Applied Colleges. As presented by table 8, Omani academic staff makes 25% of the total number of academics while non-Omanis make 75%. In addition, males make 72.5% while females make 27.5% of the total number of the staff.

Nationality/ Gender	Male	Female	Total Population by Nationality
Omani	97	63	160 (25%)
Non- Omani	367	113	480 (75%)
Total Population by Gender	464 (72.5%)	176 (27.5%)	640 (100%)

Table 8. The population of Academic Staff in Applied Colleges Regarding Gender and Nationality for 2008/2009 Academic Year

Academic staff members from the three Applied Colleges were randomly chosen to participate in this research. 12.3% of Omani staff participated in this study while non-Omani participants made 87.7%. Also 70.5% of the participants were males whereas (29.5%) were females. Table number 9 shows the staff sample regarding their nationality and gender.

Nationality/ Gender	Male	Female	Total sample by Nationality
Omani	8	7	15 *(12.3%)*
Arabic non Omani	28	4	32 *(26.2%)*
English non Omani	50	25	75 *(61.5%)*
Total sample by Gender	86 *(70.5%)*	36 *(29.5%)*	122

Table 9. The Sample of Academic Staff Regarding Gender and Nationality

Regarding the interviews, a sample of 23 academic staff from the three colleges was interviewed. Most of those interviewed were teaching foundation year students. This is because they are the most appropriate ones who are able to provide the most accurate information about these students and also because they would be the ones to apply CAA if it were to be implemented.

5.6. Research Procedures

5.6.1. Data Collection

Participants were given a brief session about Computer Assisted Assessment explaining its definition, advantages and disadvantages as well as how they will use the computer to answer the questions. All participants were asked to sign a form of consent. Then, participants were randomly divided into two groups (experimental group and control group) according to class choices. The experimental group sat the computerized test while the control group sat the paper test. Computer experience and computer self efficacy questionnaires were distributed to all the participants.

During the paper test, information was given to the participants verbally such as the restricted test duration (two hours), the need to complete the test and the awareness of plagiarism issue. Participants were given 5 minutes to complete the background questions that appear on the first page of the questionnaire (paper group) such as: Student's academic number, Gender, Region and College.

However, participants in the computer group were informed about the test laboratories and the test due time one day before. Also each participant was given a user name and a password for the Moodle programme to enable them to get access to the test. Each student's user name identified the participants' identity as well. More computer labs than needed were actually booked in advance in case there would be a problem with any lab. The main concern was to get all the participants to have the test at the same time. During the computer test, a countdown timer displaying minutes only was positioned at the top right hand corner of the computer screen. The computer automatically terminated the test session when time had elapsed.

Participants were assigned to complete the ATCAS at the end of the test session, subsequently; questionnaire booklets were collected from all participants. Before starting the questionnaire, they were asked to complete the background data that is mentioned above then write their user names and passwords in order to confirm the students' background information and identity.

Volunteer students from those allocated to take the computer test participated in the focus groups before and after sitting for the computerized test. The researcher conducted a two-hour focus group session in order to explore students' reactions and feeling towards computerized assessment. My role within the focus group was to prompt the group and encourage all the participants to be involved in the discussion, especially females who were reluctant and required a lot of encouragement in order to articulate their feelings and attitudes towards computerized assessment. As an interviewer I was also aware of the drawbacks of focus group research such as the difficulty of separating individual viewpoints from the collective viewpoint, so a lot of effort was made to fully explore the diversity of opinions within the group as well as the degree of consensus on the given topic. The participants were asked about their prior experience in computers and computerised tests as well as their opinions about switching to computerised assessment. None of them reported having had any experience in computerised testing. Also, they were asked about the difficulties they encountered in the computerized test and whether they would prefer to be assessed in Arabic instead of English and the reasons behind that. However, the same questions were given to the participants before and after taking the computerized test in order to trace any change in their feelings and find out whether or not their beliefs and thoughts have changed and in what direction their attitudes may have changed and why.

In regard to academic staff, all participants were asked to sign a form of consent They were assigned to complete a questionnaire in order to investigate their attitude(s) towards CAA Then semi-structured interviews for the staff were conducted to explore in depth their thoughts and opinions about CAA The staff interviews took place in each college and were carried out in each member of staff's office. Each interview took 30 minutes. The interviews were recorded so that they would be analysed later in depth and also to make the interviewee feel relaxed during the interview as recommended by Burton & Bartlett (2005). These experimental procedures are summarized in the chart (Figure 10) below.

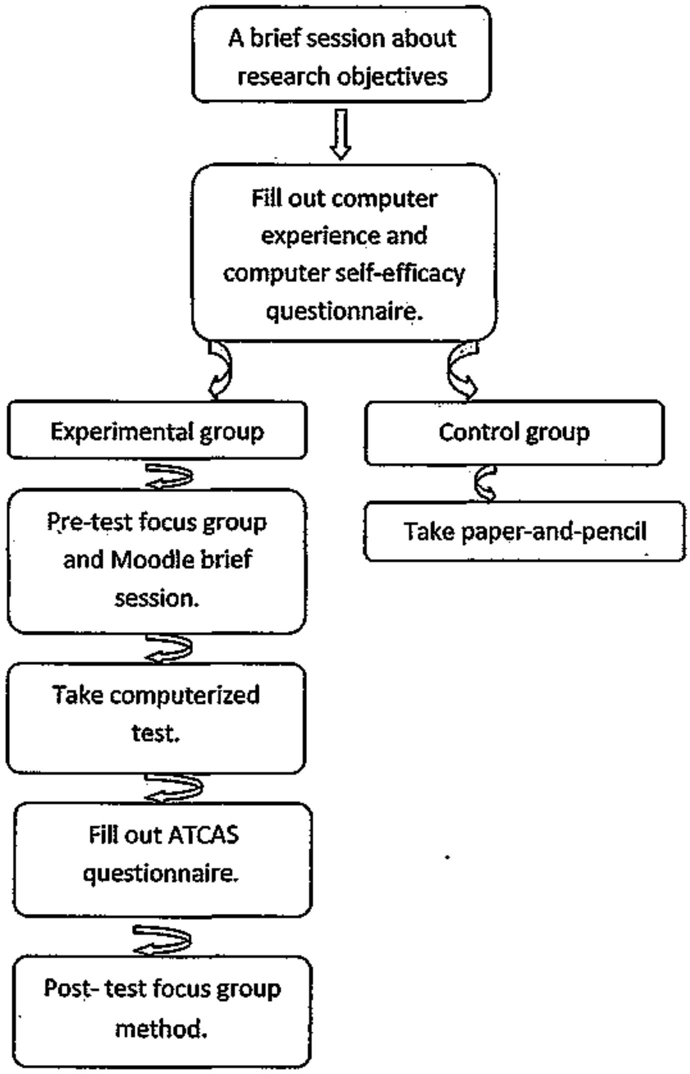


Figure 10. Summary of Experimental Procedure for Students

5.6.2. Data Analysis

As this study used multiple methods the data analysis was varied. The quantitatively obtained data were analysed statistically. Firstly, graphs and descriptive statistics were obtained to indicate the general view about data. In addition, in order to compare different groups, an independent t-test was conducted to variables contain two groups and one way ANOVA for variables with more than two groups. Then Multifactor ANOVA was conducted. Covariate ANOVA was used to detect the relationship between some variables. In addition factor analysis using a principal component was done in the validation study and main study as well. Reliability (Cronbach's alpha coefficients) was calculated for all study instruments. The Statistical Package for the Social Science for Windows (SPSS) was used to perform the statistical analysis.

The qualitative data obtained from the focus groups and the semi- structured interviews was analysed thematically through a thematic content analysis. This method was used to identify participants' views, experiences, feelings and thoughts towards CAA according to the obtained data. The audiotapes were then transcribed and the content of transcripts was thoroughly studied The material collected was later reduced by selecting, focusing, simplifying, abstracting and transforming the raw data (Miles & Huberman, 1994).

Using multiple methods of data collection allows triangulation in this study which enhanced the trustworthiness of the findings. For example the consistency of the data obtained from questionnaire and confirmed by interviews or focus groups will be source of validity evidence and will increase the credibility and dependability of the study finding.

5.7. Ethical Considerations

Ethical Approval was obtained from the University of Plymouth and the Ministry Of Higher Education in Omán to use tests and questionnaires and conduct group interviews for students as well as interview teachers and distribute a questionnaire. All the details of the study were included in a covering letter and participants, students and teachers, signed a consent form before participating in the study (See Appendix A). The consent form contained a description of the purpose of the study, details of the data gathering methods, a description of the potential benefits of the research and an assurance that participants could withdraw from the study at any time without penalty. Furthermore, before starting the data collection phase, the deans and heads of the English department of each college were informed about the aims of the study, the research timetable and the proposed data gathering techniques.

All the data was treated in a way which protected the confidentiality and anonymity of the students and teachers involved in the study. Students were informed that each could get their test results by email on request. Moreover, students and teachers were informed that their identities would remain anonymous, and that no personal information along with the signed consent forms would be disclosed either verbally or in the publications of any document associated with project.

At the end, I have got good instruments that have enabled me to get reliable and trustworthy data, which helped me to fit together the whole picture. The next chapters will be devoted to the analysis and discussion of all the data that was obtained from many sources.

Chapter Six: Exploring the Equivalence of the Students Performance

6.1. Introduction

This research seeks to investigate the equivalence of both assessment modes (traditional paper-based and computer-based assessment). This chapter is the most important in the thesis because it addresses the key question of whether student test scores are affected by the mode of administration It also investigates whether or not there are any differences in performance for each mode of delivery for students of different gender, from different colleges or different regions of residence.

First of all, the sample characteristics are described to give a general overview. Then, the preliminary analysis of test scores is made, using graphical displays and descriptive statistics in order to provide a general view of the whole data. A sequence of increasingly complex statistical analyses was performed. First, mean scores were compared in terms of groups (by mode, gender, college and region) ignoring all the others factors. Then, all grouping factors were combined into a single analysis (Multi-Factor Analysis of Variance) to determine the effect of each factor when all others are taken into account simultaneously.

There are several advantages for using Multi-Factor Analysis of Variance such as reducing the risk of type 1 errors which separate ANOVAs can cause through multiple testing (Pallant, 2005). Moreover, unequal numbers between subgroups may cause bias in the one way analysis result. for example, if females outnumber males in one college only, any gender difference may appear as a difference between colleges. Hence, Multi-Factor ANOVA is very useful, because it can help us identify any significant main effects or interactions. Finally, all the findings of the chapter will be summarised and discussed in the light of the literature.

6.2. Sample Characteristics, Preliminary Analyses and Descriptive Statistics

6.2.1. Sample Characteristic

Table 10 below shows the sample sizes and sample summary statistics for students' performance in relation to mode, gender, college and region.

Variables		Count	Percentage	Mean	SD
Mode	*Paper*	*243*	*55.4%*	*20.55*	*6.95*
	Computer	*196*	*44.6%*	*20.23*	*6.41*
Gender	*Male*	*227*	*51.7%*	*16.99*	*5.31*
	Female	*212*	*48.3%*	*24.08*	*6.09*
College	*Sur*	*133*	*30.3%*	*18.08*	*5.56*
	Ibri	*163*	*37.1%*	*21.28*	*6.66*
	Nizwa	*143*	*32.6%*	*21.58*	*7.23*
Region	*Alsharqiya*	*175*	*39.9%*	*18.82*	*6.11*
	AlBatinah	*88*	*20%*	*21.82*	*6.69*
	AlDhahira	*117*	*26.7%*	*20.57*	*6.96*
	AlDakhlia	*59*	*13.4%*	*22.71*	*6.93*
Total students sample, Mean and SD		*439*	*100%*	*20.41*	*6.7*

Table 10. Demographic Data of Respondents and the Total English Test Scores Mean and SD According to Mode, Gender, College and Region. Scores are out of 50.

Of the 439 students assigned to the paper and computer test, 227 (51.7%) were males and 212 (48.3%) were females. According to the class allocation, 243 (55.4%) of the respondents sat for the paper-based test and 196 (44.6%) sat for the computerized one. This cohort comprises students belonging to three

Applied Science Colleges, 133 students in Sur, 163 in Ibri and 143 in Nizwa which respectively make up 30.3%, 37.1% and 32.6% of the total student number. Once again we can note that these students come from different regions of the Sultanate as 175 students (39.9%) come from AlSharqiya, 117 students (26.7%) come from AlDhahira, 88 students (20%) from AlBatinah and 59 (13.4%) come from AlDakhilia region .

6.2.2. Preliminary Analysis and Data Summary

Prior to conducting the main analysis of performance, inspection of the graphical and numerical summaries in the Appendix B, indicates that scores could be considered to be normally distributed for the total English language scores and its components (grammar and reading). The purpose of this section is to give a general overview of the data before presenting the result of formal hypotheses tests.

The table 10 above shows the total English test scores according to mode, gender, college and region. It seems that the means of the total English test scores are very similar across modes of administration.

The results obtained also indicate that in total, female students seem to have scored higher than male students in the total English test. As for differences between colleges, it can be seen that while there are similar means for the students at Ibri and Nizwa Colleges, the performance of the students at Sur College was somewhat lower.

In regard to differences between students from different regions, there are almost equal mean scores between AlDhahira, AlBatinah and AlDakhlia regions whose mean scores were higher than students from AlSharqiya. This is expected because the majority of AlSharqiya students are studying at Sur

College which also had the lower mean compared to other colleges. In addition, the standard deviations (SD) reflect almost equal variation between students from different regions.

Table 11 below shows the mean and the standard deviation (SD) of the grammar and the reading tests in relation to the mode of delivery, gender, college and region.

Variables		**Grammar**		**Reading**	
		Mean	*SD*	*Mean*	*SD*
Mode	*Paper*	*12.79*	*4.71*	*7.76*	*3.07*
	Computer	*12.76*	*5.06*	*7.47*	*2.87*
Gender	*Male*	*10.54*	*3.96*	*6.44*	*2.53*
	Female	*15.17*	*4.61*	*8.90*	*2.90*
College	*Sur*	*11.54*	*3.94*	*6.54*	*2.76*
	Ibri	*13.25*	*5.29*	*8.03*	*2.89*
	Nizwa	*13.39*	*5.29*	*8.19*	*3.04*
Region	*AlSharqiya*	*11.92*	*4.35*	*6.89*	*2.83*
	AlBatinah	*13.63*	*4.98*	*8.19*	*2.83*
	AlDhahira	*12.63*	*5.24*	*7.94*	*3.11*
	AlDakhlia	*14.32*	*4.90*	*8.38*	*2.98*
Mean and SD for the whole sample		*12.77*	*4.87*	*7.63*	*2.98*

Table 11. Grammar and Reading Mean and SD in relation to (Mode, Gender, College and Region). Grammar Scores are out of 30. Reading scores are out of 20.

The grammar mean scores for both modes are similar. Again females have a higher mean score than males. In addition, there is dissimilarity of means regarding colleges: Sur College has a lower mean score compared to Nizwa and Ibri which seem to have a similar means. Regarding the region variable, AlDakhlia has higher mean scores than the other three regions. Whereas AlSharqiya region has the lowest mean comparing to the rest of regions

The table also shows that both paper and computer delivery modes give similar means regarding the reading test Yet again, females seem to have slightly higher mean scores than males. In addition, there are differences regarding colleges: Sur College has a lower mean score compared to Nizwa and Ibri which seems to have *similar* means. Regarding to the region, we can see that AlDakhlia and AlBatinah seem to have a similar means, whereas Al Dhahira has a slightly higher mean than AlSharqiya.

6.3. Test of Equality of Means

The previous section suggests that there may be a difference in students' performance in relation to gender, college and region variables but not mode. Undertaking more rigorous formal statistical analysis, we should note that the first criterion for demonstrating test equivalence is the equality of means. Therefore, this section will compare the scores from different groups using an independent t-test and one way ANOVA to determine whether there is a statistically significant difference between the mean test scores in relation to group variables such as mode, gender, college, and region of residence.

6.3.1. Effect of Administration Mode on the Total English Language Test

In order to test the hypothesis which is *H0 1:There is no difference between average test scores in relation to the method of delivery* , looking at the box plots in figure 11 indicates the total test mean is almost equal between paper and computer groups.

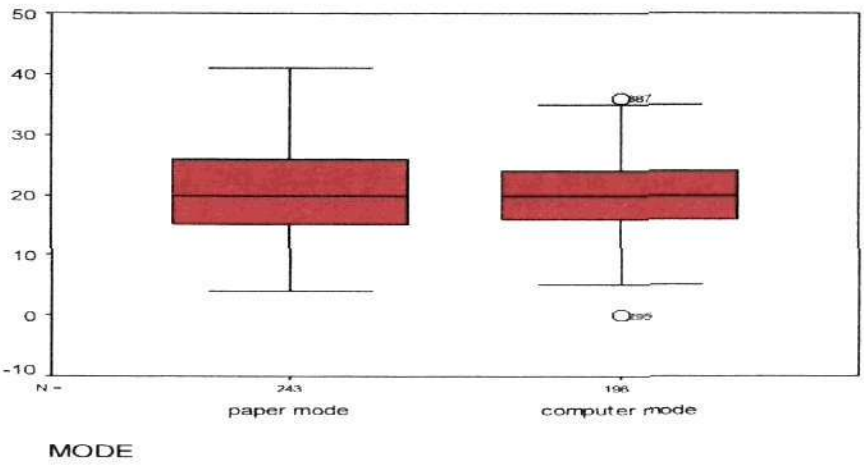


Figure 11. Mode Box Plots for the Total English Test Score

An independent t-test was first performed to determine whether both groups yielded comparable mean results of the English language test or not (Appendix B). This test was found to be not statistically significant [t (437) =0.504, p=0.615], indicating the equivalence in average student performance between those who have taken the paper-based and those taking the computer-based tests.

In addition, the equality of variances was tested using Levene's test for equality of variances in independent samples. The result of the test (Appendix B) indicates that the paper and computer tests did not differ significantly in their variances (*F=2.677, p=0.103)*.

The difference in means between the paper-based and the computer based administration groups was also evaluated using the effect size calculated by the eta squared index (Pallant, 2005). The effect size was very small (0.0005) according to Cohen's (1988) criterion. That means that there is no practical difference, or only 0.05 percent of the variance in total English test score was explained by mode.

6.3.2. Effect of Administration Mode on the Grammar and Reading

Regarding grammar and reading, the preliminary analysis, using box plots in the figure 12 and 13, indicates the grammar means are roughly comparable between the paper and the computer groups. However, the reading test has a higher mean score for the paper test compared to the computer one.

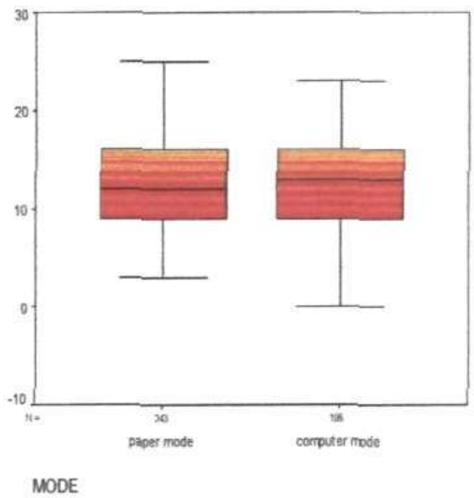


Figure 12. *Mode Box Plots for Grammar Test*

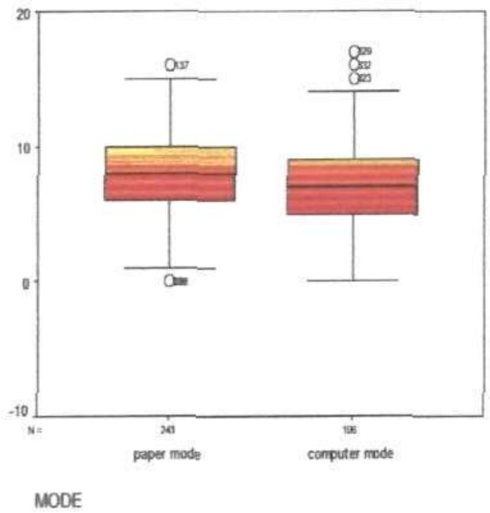


Figure 13. *Mode Box Plots for Reading Test*

The t-test result was not statistically significant *[t (437)= 0.073, ρ= 0.942];* [t(437)=1.01,ρ=0.311] for both grammar and reading respectively, indicating that the paper and the computer groups yielded comparable mean scores in both grammar and reading tests. Moreover, the result of the Levene's test indicated that the paper and the computer test did not differ significantly in their variances (*F=0.625; 0.313, ρ=0.430; 0.576*) (See Appendix B).

The effect size (Eta Squared) of the mean difference was also calculated. Based on Cohen's standards (1988) the effect size for both grammar (0.000005) and reading (0.0023) were extremely small. In other words, there is almost no variance in grammar and reading which could explain by mode.

This finding verified that students in general did not seem to be affected by the mode of delivery and their performance was related only to the test content.

6.3.3. Gender Variation on the Total English Language Test Score

In order to test the hypotheses which is *Ho: There is no difference between average test scores and its parts (grammar and reading) in relation to examinee's gender;* a t-test was conducted to investigate the gender difference.

The first glance at the box plots (Figure 14) indicates that the mean total English test scores were not comparable between males and females.

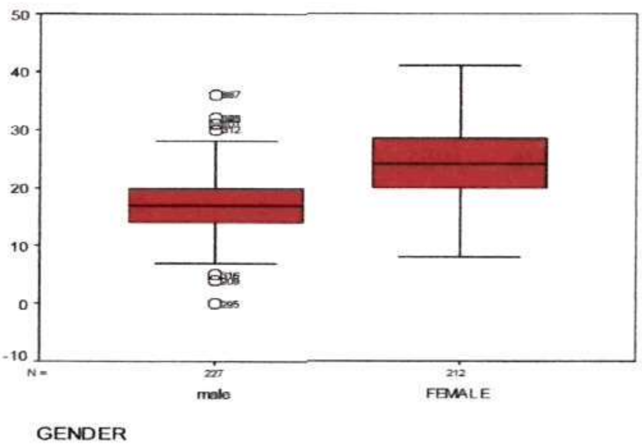


Figure 14. *Gender Box Plots for the Total English Test*

The t-test shows that females performed significantly higher than males in the total English test scores [$t\ (439) = 12.941$, $p < 0.0005$]. Females also show a slightly greater variance than males. Since the assumption of equal variances was violated in the total test score [Levene's test $F=8.86, \rho=0.003$] , the result presented is based on an alternative t-value which compensates for the fact that the variances are not the same (See Appendix B). The magnitude of the difference in the means was large (Eta Squared=0.276) in terms of Cohen's (1988) criterion. It is not clear what the cause of this difference might be.

6.3.4. Gender Variation on the Grammar and Reading Test

Preliminary analysis using box plots (Figure15 and 16) indicates that the females also performed better than males in both the grammar and the reading components of the test.

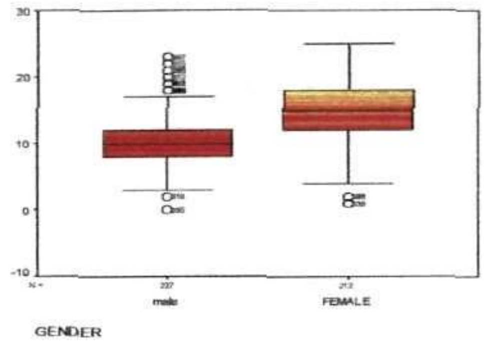


Figure 15. *Gender Box Plots for Grammar Test*

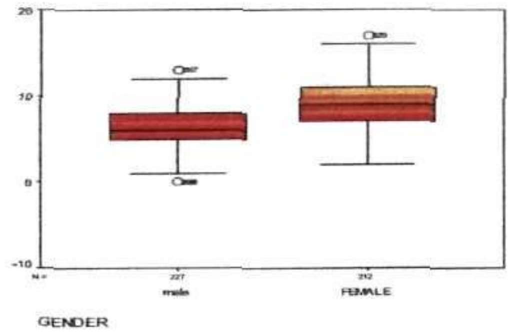


Figure 16. *Gender Box Plots for Reading Test*

The t-test results (Grammar [$t(439)=11.31$; $p<0.0005$; Reading [$t(439)=9.42$; $p< 0.0005$)] show a significant difference in terms of gender with females achieving higher scores than males in both grammar and reading (See Appendix B). In addition, the effect size (Eta Squared) was large for both grammar (0.226) and reading (0.166) according to Cohen's criterion.

One question which arises is whether males and females performed differently across the two modes of administration. This is exactly the sort of question that the Multi-Factor Analysis described later is designed to answer.

6.3.5. Colleges Variation of the Total English Test Score

Looking at the figure 17 indicates that Sur College has the lower mean scores than the other two colleges which appear almost comparable.

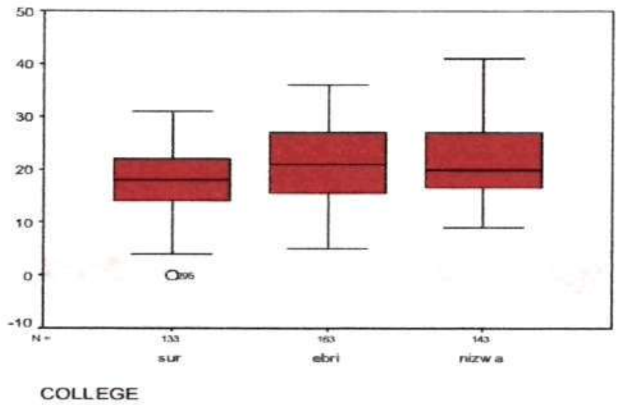


Figure 17. Colleges Box Plots for the Total English Test

One way ANOVA, presented in the Appendix B, showed a statistically significant difference between colleges in the mean total test scores [$F(2,436)=12.17$, $p< 0.0005$]. Despite reaching statistical significance, the actual difference in mean scores between the groups was quite small according to Cohen's (1988) terms. The effect size calculated using Eta squared was 0.053.

Follow up analysis (Post Hoc Comparison) using the Tukey HSD test indicated that the lowest mean score for Sur College (x=18.08,SD=5.56) was significantly different from Nizwa and Ibri colleges [Ibri (x= 21.28, SD =6.66) and Nizwa (x= 21.58, SD=7.23], while Nizwa and Ibri means did not differ significantly from each other (See Appendix B).

So, could the College performance difference be related to the difference in composition of student body, or difference in proficiency on entry to college, or difference in teaching?

6.3.6. College Variation of the Grammar and Reading Test.

Preliminary analysis (Figure 18 and 19) also indicates that Sur College has *lower mean scores than the other two colleges in both grammar and reading.*

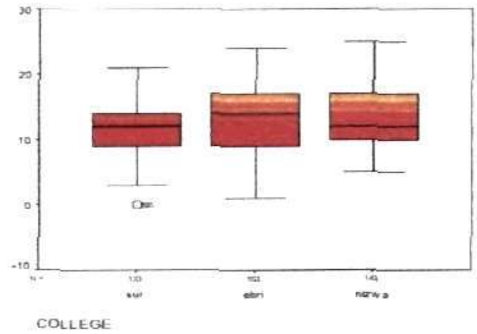


Figure 18. Colleges Box Plots for Grammar Test

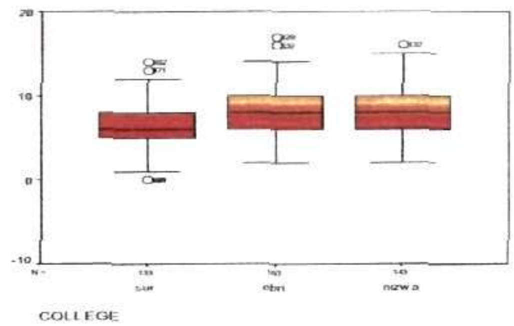


Figure 19. Colleges Box Plots for Reading Test

Hence, one way ANOVA was conducted to explore if there is a significant differences in students' performance over the three colleges (Sur, Ibri and Nizwa). There was a statistically significant differences in grammar and reading $F(2,436)=6.34,\ p=0.002]$; $[(F(2,436)=13.63,\ p< 0.0005)]$ respectively (See Appendix B). Despite reaching statistical significance, the actual difference in mean scores between the groups was small in the grammar (0.028) and reading (0.059) according to Cohen's (1988) terms.

Follow up analysis (Post Hoc Comparison) using the Tukey HSD test indicated that the mean score in grammar and reading for Sur College was significantly different from Nizwa College and Ibri College, while the means of Nizwa and Ibri colleges did not differ significantly from each other (Appendix B).

6.3.7. Regional Variation on the Total English Test

Preliminary analysis (Figure 20) indicated that students coming from AlDakhlia have the highest mean scores whereas AlSharqiya students have the lowest. Moreover, it seems that AlSharqiya and Al Dhahira were almost comparable.

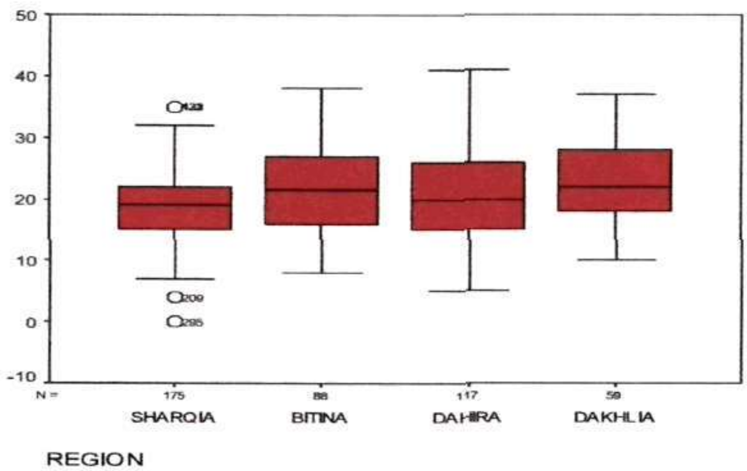


Figure 20. Regions Box Plots for the Total English Test

One way ANOVA was conducted to explore if there is a significant difference in students' performance over the regions (See Appendix B). There was a statistically significant mean effect between regions in the total English test scores *[F(3,435)=7.19, p< 0.0005]*.

Follow up analysis (Post Hoc Comparison) using the Tukey HSD test indicated that there was no significant mean difference between AlDhahria region and the other three regions. In addition, the lowest mean score for AlSharqiya region was significantly different from AlBatinah and AlDakhlia regions (Appendix B).

However, despite reaching statistical significance, the actual difference in mean scores between the groups was quite small according to Cohen's (1988) terms. The effect size calculated using eta squared was (0.047). It is important to note that most Alsharqiya region's students are studying at Sur College which is also has the lowest performance compared to the other two colleges.

6.3.8. Regional Variation on the Grammar and Reading

Looking at the box plots (Figure 21) of the grammar component indicates that *AlDakhlia students have the highest mean scores whereas AlSharqiya students* had the lowest. But looking at reading box plots (Figure 22) it seems that all three regions except AlSharqiya had similar means.

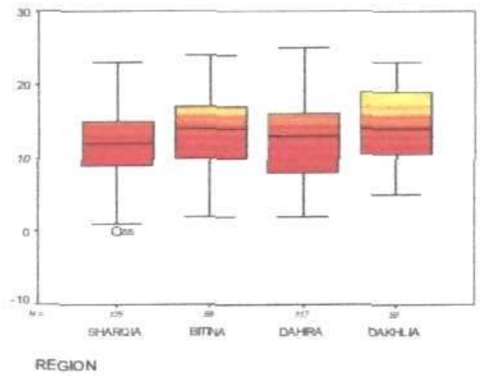


Figure 21. Regions Box Plots for Grammar Test

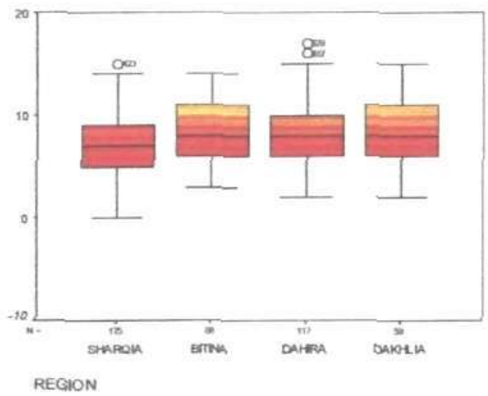


Figure 22. Regions Box Plots for Reading Test

There was a statistically significant effect between region means in grammar *[F(3,435)=4.83, p=0.003]* and reading *[F(3,435)=6.47, p< 0.0005]* (See Appendix B).

Follow up analysis (Post Hoc Comparison) using the Tukey HSD for grammar test indicated that there was no significant mean difference between AlDhahria region and the other three regions. In addition, the lowest mean score for AlSharqiya region was significantly different from AlBatinah and AlDakhlia regions (See Appendix B). Similarly, in the reading test; it has been found that AlSharqiya lower mean score was significantly different from the other three regions. Whereas the other three regions mean scores did not differ significantly (See Appendix B).

Despite reaching statistical significance, the actual difference in mean scores between the groups was quite small according to Cohen's (1988) terms. The effect size calculated using eta squared was 0.032 for grammar and 0.043 for the reading.

6.4. Test of Equality of Means Using Multi-Factor Analysis of Variance

These preliminary analyses suggest that, ignoring factors such as gender or region of origin, the mode of delivery does not affect test scores. However, the existence of significant gender, college and regional differences leads to considering the possibility of the existence of the question of whether there may be a mode effect which differentially influences scores for males and females. Accordingly, a Multi Factor ANOVA was undertaken. In this method, overall effects such as the difference between genders mean show up as main effects. Any differential effects of mode between the genders would show up as a significant interaction (denoted MODE*GENDER) in SPSS output. As well as the differences between gender, college and regions seen previously, this more complex thorough and sensitive analysis now (Table 12) shows a small but significant effect of mode, and a highly significant interaction between mode and gender. This different pattern arises because the comparison between modes (for example) takes the difference between genders into account, rather than *ignores it*

Tests of Between-Subjects Effects

Dependent Variable: TOTAL

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Eta Squared
Corrected Model	8305 388[a]	35	237 297	8.373	000	.421
Intercept	101238 153	1	101238 153	3572 282	000	899
MODE	*127 578*	1	*127 578*	*4 502*	*034*	*011*
GENDER	1817.524	1	1817.524	64 133	000	137
COLLEGE	215.408	2	107.704	3 800	023	019
REGION	240.325	3	80.108	2.827	.038	021
MODE * GENDER	339.090	1	339.090	11 965	001	029
*MODE * COLLEGE*	*10 345*	2	*5 172*	*183*	*833*	*001*
GENDER * COLLEGE	16 185	2	8 093	286	752	001
MODE * REGION	35 681	3	11.894	.420	739	003
GENDER * REGION	205 091	3	68.364	2 412	066	018
COLLEGE * REGION	179 298	3	59 755	2 109	099	015
Error	11420 986	403	28.340			
Total	202641.000	439				
Corrected Total	19726.374	438				

a R Squared = 421 (Adjusted R Squared = 371)

Table 12. Multi Factor Analysis of Variance for the Total English Test Score

When all *explanatory* factors are examined together, a different picture emerges. Mode now shows a significant effect on the total test score which is different from the conclusion based on just a simple t-test (section 6.3.1.). This result indicated that students performed better in the paper test than in the computer one, even though the difference in mean was less than one point (0.32). Also It should be noted that the effect size was small but significant for the total test (Eta Squared=.011)

However, no mode difference was detected in the grammar and reading tests (Table 13, 14). Also the effect size was small for grammar and reading respectively (Eta squared= 0.006, 0.009).

Tests of Between-Subjects Effects

Dependent Variable: TOTGRAMM

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Eta Squared
Corrected Model	3670.159a	35	104.862	6.285	.000	.353
Intercept	39278.013	1	39278.013	2354.123	.000	.854
MODE	39.209	1	39.209	2.350	.126	.006
GENDER	796.728	1	796.728	47.752	.000	.106
COLLEGE	57.112	2	28.556	1.712	.182	.008
REGION	116.121	3	38.707	2.320	.075	.017
MODE * GENDER	325.570	1	325.570	19.513	.000	.046
MODE * COLLEGE	2.380	2	1.190	.071	.931	.000
GENDER * COLLEGE	1.035	2	.517	.031	.969	.000
Error	6723.964	403	16.685			
Total	82059.000	439				
Corrected Total	10394.123	438				

a. R Squared = .353 (Adjusted R Squared = .297)

Table 13. Multi Factor Analysis of Variance for Grammar Test

Tests of Between-Subjects Effects

Dependent Variable: TOTREAD

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Eta Squared
Corrected Model	1168.060a	35	33.373	4.898	.000	.298
Intercept	14397.248	1	14397.248	2113.213	.000	.840
MODE	25.293	1	25.293	3.713	.055	.009
GENDER	208.360	1	208.360	30.583	.000	.071
COLLEGE	54.413	2	27.207	3.993	.019	.019
REGION	22.462	3	7.487	1.099	.349	.008
MODE * GENDER	.160	1	.160	.023	.878	.000
MODE * COLLEGE	5.787	2	2.893	.425	.654	.002
GENDER * COLLEGE	14.299	2	7.150	1.049	.351	.005
COLLEGE * REGION	30.422	3	10.141	1.488	.217	.011
Error	2745.626	403	6.813			
Total	29508.000	439				
Corrected Total	3913.686	438				

a. R Squared = .298 (Adjusted R Squared = .238)

Table 14. Multi Factor Analysis of Variance for Reading Test

Gender shows a highly significant effect on the total test, grammar and reading scores, which confirmed the preliminary analysis using t-test. The effect size was also large for the total test and moderate for grammar and reading (Eta Squared=0.14, 0.11, 0.07).

The college main effect was significant for the total test score and for reading which confirmed the result by one way ANOVA analysis as well. However, the grammar test shows no significant difference over the three colleges, which is different from the conclusion based on one way ANOVA. The effect size was small for total test, grammar and reading (Eta squared= 0.019, 0.008, 0.019).

The region main effect was significant in the total test score which also confirmed the one way ANOVA result. Whereas no significant difference was detected in the grammar and reading across the different regions, which is different from the conclusion based on one way ANOVA.

The only significant interaction was between gender and mode of assessment and the mean plot below (Figure 23) clarifies that females achieved better scores in the paper while males achieved better in the computer one.

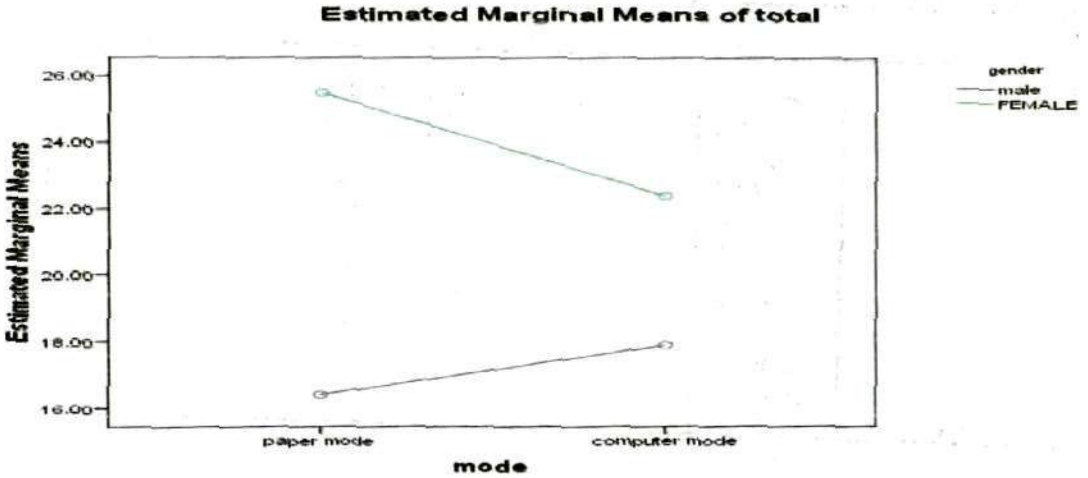


Figure 23. Estimated Marginal Means of the Total English Test Score

As the number students in each group is not equal, estimated marginal means are used to compare the means of unequal sample sizes which take proportion of the sample size into consideration. Table 15 below shows the total English test score estimated marginal means and standard error for male and female performance in both modes separately.

Mode	Male M (Std.E)	Female M (Std.E)	Mean of Mode
Paper	16.33 (0.486)	25.58 (0.530)	20.55
Computer	17.91 (0.573)	22.43 (0.556)	20.23
Mean of Gender	16.99	24.08	20.41

Table 15. Estimated Marginal Means and Standard Error for the Total English Test Performance between Gender/Mode

Further exploration of the gender/mode interaction using t-test, presented in Appendix B, showed a significant mean difference which shows that males tend to achieve better in the computer test compared to the paper test of the total English test scores *[t (227)=2.116, ρ=0.036*]. Moreover, the magnitude of the difference in the means was large (Eta Squared=0.276) in terms of Cohen's (1988) criterions.

In respect to females, there was a significant difference as well but, their achievement in the total test scores was higher in the paper test than in computer test [t (212) =3.88, p < 0.0005). Moreover, the effect size was moderate (Eta Squared= 0.07) according to Cohen's (1988) terms (Appendix B).

In regard to (Mode*Gender) grammar test, there was a significant mean differences for males as showed by t-test *[t (227) =2.57, ρ=0.01)]*, where males did better in the computer test *(x=11.38, SD=4.78)* than in the paper test

(x=9 93, *SD=3.12*). Unlike males, females achieved higher scores in paper (*x=16.19, SD=3.97*) compared to computer (*x=14.06, SD=5.00*) with significant means differences between means *[t (212) =3.44, p<0.0005)]* (Appendix B).

In contrast, males in the reading test component showed no significant difference across modes [*t (227) = 0.387, p= 0.699*)] which means that males yielded comparable scores in the computer test (x=6.53, SD=2.45) and the paper one (x=6.39, SD=2.60). However, females students still achieved better in paper test (x=9.40, SD=2.80) than in computer one (x=8.37, SD=2.96) with a significant difference between means *[t (212) =2.60, p=0.010)]* (Appendix B).

These results indicate that males may benefit more from computer tests while females may be disadvantaged by them. This fact was hidden in the earlier simple analysis and has only shown up because a more sensitive statistical approach has been used. There are many possible explanations for this result, for example the potential effect of computer experience or computer self efficacy differences, or also perhaps the students attitudes' towards computerized test. It might also be explained by computer anxiety and preference differences between males and females. The next chapters will explore these possibilities.

6.5. Discussion

The findings of the present study show a small mode difference in the total English test score with students tending to perform slightly higher in the paper test compared to the computerized one. It is interesting that the mode difference is very small (0.32), which can be explained by the fact that as students used to the traditional test so they got higher score on paper test.

Another explanation might be brought by the effect of other variables such as computer experience or computer attitudes which will be investigated later. Also it might be explained as Dimock & Cormier, (1991) suggest that computerized testing on the first occasion may produce lower scores, and which they called the "novelty factor". Also Fulcher (1999) asserts that there is a strong rapport between familiarity with computers and students' achievement on computerised tests and he stresses the importance of ensuring that students are familiar with the testing mode, for assessment results would otherwise be impacted by other variables than the test content. According to Fulcher, "the issue of familiarity is not new in language testing. It has always been accepted that test-takers should be familiar with the item types and mode of test delivery before taking a test to ensure that these factors would not be confounding variables in score interpretation" (1999, p.291). Accordingly, lack of computer familiarity may translate into impaired student performance on computer-based assessment.

It may also be explained as Wolfe & Manalo (2004) argue that "it is likely that such an effect would be more pronounced for [the] examinees for whom English is a second language because these examinees would perform a double translation – native language to English and then English to keyboard strokes" (p.54).

The most interesting finding is that although females generally score higher than males in both modes, females score lower in computerized test than in paper tests while males score higher on computerized rather than in paper tests. It is hard to explain this result as for both males and females this was the first encounter to the computerised assessment but males' performance was higher on computer which seems that the novelty of mode did not affect them as females did. However, one possible explanation might explain this result, is the fact that in Oman, boys have access to the computers more than girls inside or outside homes and schools (see Chapter Three).

Wallace & Clariana (2005) verified that females tend to have high scores when they have the same amount of practice on computers as males In their study, at the beginning, females had lower scores than males before the course. They also note that most studies have found that males use computers more often than females inside and outside homes. However after taking the same extensive fundamental computer course, females tended to have higher scores than males. This means that when females are given the same opportunities to deal with computers as males then females will not be disadvantaged by computerised tests.

Wolfe & Manalo (2004) have also investigated the impact of the test mode effect on the performance of learners of English as a second or a foreign language in a direct writing assessment. They have found out that those with a higher English language proficiency level performed better in both testing modes while those whose, level of English is rather low, performed better in the paper-based test. They state "we found a weak two-way interaction between composition medium and English language proficiency with examinees with weaker English language scores performing better on handwritten essays while examinees with better English language scores performing comparably on the two testing modes" (Wolfe & Manalo, 2004, p.53). However, in the present study females seem to have higher English proficiency level, as their performance was higher than males in both modes, but they performed better in the paper test compared to computer one, while males preformed better in the computer based test. So, this simple explanation may not be correct here. This interaction between gender and mode may suggest searching deeper into psychological processes and attitudes to identify the important factors that affect female's performance in computerised assessment.

There was a difference in the total scores among colleges and regions. Sur College had the lowest score compared to the others. Most students who study at Sur College are originally from AlSharqiya region which also had the lowest scores compared to the other three regions. As other colleges had students from different regions and Sur College has students only from one region (AlSharqiya) that might reduce the students' motivation and competition. It is difficult to explain the potential cause of this finding, but this may relate to the students' computer experience or their attitudes towards computers, which may affect students' performance. That is what the next chapters are intended to explore.

Chapter Seven: Students' Computer Experience and Computer Self-Efficacy

7.1. Introduction

It has been shown in the previous chapter that males and females perform differently depending on the mode of test administration. Also there were small but significant differences in relation to the mode, college and region in terms of the total English test score. So, the purpose of this chapter is to investigate other factors which may explain these differences. In particular, I have examined whether computer experience and computer self efficacy differ in terms of gender, college or region and whether the level of computer experience and computer self efficacy can partly explain the differences in test scores.

The same statistical approaches (procedures) will be used through this chapter as in the previous one. Even though the study instruments were evaluated during the validation study, the results will be confirmed on this larger sample. Finally, a discussion will summarise the findings of this chapter in the light of the literature.

7.2. Brief Description of Data Collecting and Organisation

7.2.1. Computer Experience Questionnaire

The computer experience questionnaire, presented in Appendix A, consists of eight items taken from many studies (Al-Kother, 1999; Johnson, Ferguson & Lester, 1999, Smith & Caputi, 2004) which intend to measure the amount and kind of students' computer experience. All eight items were coded in order to let higher marks reflect high computer experience. The overall marks were summed giving a score between 8 to 24 marks: 8 expresses very little computer experience while 24 expresses lots computer experience.

7.2.2. Computer Self-Efficacy Questionnaire

Seven items from the computer self efficacy scale originally developed by Fagan, Neill & Wooldridge (2004) were used. The individual marks were summed to give a score between 7 and 35 marks: 7 expresses very low computer self efficacy and 35 expresses very high computer self efficacy.

7.3. Factor Analysis Validation for Computer Experience and Computer Self-Efficacy Questionnaires

7.3.1. Factor Analysis for Computer Experience Questionnaire

439 foundation year students participated in completing the questionnaire. The sample consisted of 227 (52%) males and 212 (48%) females. The eight items were subjected to principal components analysis. The Kaiser-Meyer-Oklin value was 0.76 exceeding the recommended value 0.60 (Kaiser, 1974) and the Barlett's test of sphericity reached statistical significance supporting the factorability of the correlation matrix.

Principal components analysis revealed the presence of three components with eigenvalues exceeding 1 explaining 32.7 percent, 13.5 and 12.7 of the variance respectively. The total variance explained by these three components was 59 percent.

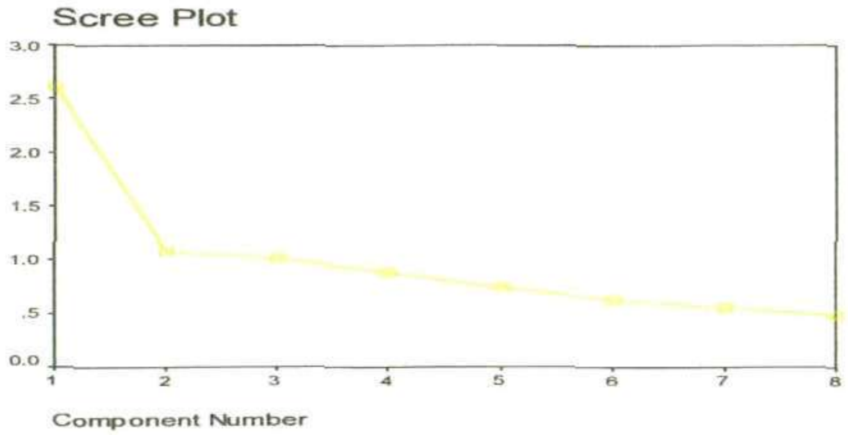


Figure 24. Scree Plot of Computer Experience

The inspection of the above scree plot (Figure 24) revealed a clear break after the first component. It was decided to retain one component for further investigation. The one component solution explained a total of 32.7 percent of the variance (See Appendix C).

Most items loaded in this component with a coefficient of more than 0.40 except one item (item number 7) which had a low coefficient (0.12). This item had a low correlation in the pilot study and this large sample confirms the finding. The internal consistency of the scale (Cronbach's alpha) was 0.67 which was the same as in the validation study. However, the alpha when item seven was deleted increased to 0.71. Because of the low loading and increase in the Cronbach's alpha, seven out of the eight items on the computer experience questionnaire were retained for further analysis.

7.3.2. Factor analysis for Computer Self -Efficacy Questionnaire

The seven items were subjected to principal component analysis. The Kaiser-Meyer-Oklin value was 0.88 exceeding the recommended value 0.60 (Kaiser,

1974) and the Barlett's test of sphericity reached statistical significance supporting the factorability of the correlation matrix.

Principal components analysis revealed that all seven items loaded on one factor with eigenvalue 4.18. The total variance explained was 60 percent (Appendix C). Moreover the scree plot (Figure 25) revealed a clear break after the first component.

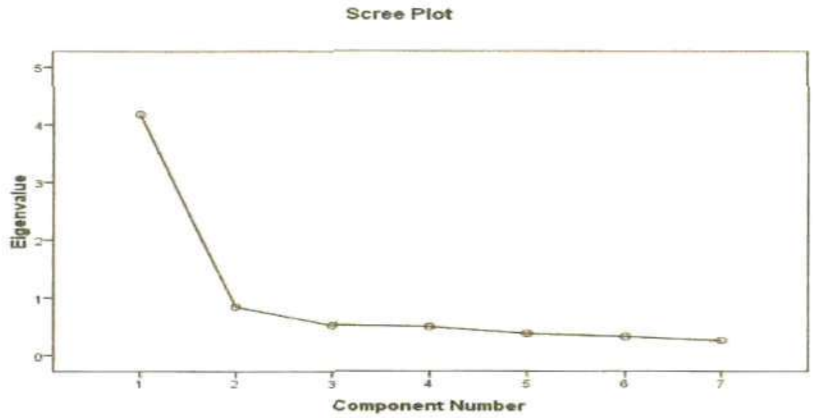


Figure 25. Scree Plot of Computer Self-Efficacy

The correlation for all items loaded in this component was 0.71 to 0.83. Also the internal consistency of the scale (Cronbach's alpha=0.89) was good. This result agrees with the internal consistency of the scale in the Fagan, Neill & Wooldridge (2004) study which was (Cronbach's alpha) 0.93 and in my validation study (Cronbach's alpha=0.93). Together these findings provide further evidence for the internal consistency of the computer self efficacy questionnaire with no modification.

7.4. Reliability Coefficient of the Instruments (Computer Experience and Computer Self-Efficacy).

Reliability coefficients for computer experience and computer self efficacy are reported in Table 16 for the entire test taking sample and separately for the paper and computer groups.

Reliability coefficient	Computer experience	Computer self efficacy
Entire sample	0.71	0 89
Paper group	0 68	0 90
Computer group	0 73	0 88

Table 16. *Reliability Coefficient for the Computer Experience and Computer Self-Efficacy*

For the total sample, reliability coefficient for computer experience and computer self efficacy was 0.71 and 0 89 respectively These coefficients indicate satisfactory internal consistencies for computer experience and computer self efficacy scales in both groups.

7.5. Analysis and Comparisons between Groups

7.5.1. Preliminary Analysis

Prior to conducting the main analysis, the inspection of the histograms in the Appendix C, and box plots indicated that total scale scores were normally distributed for both computer experience and computer self efficacy.

7.5.2. Descriptive Statistics

Table 17 below shows the mean and standard deviation for computer experience and computer self-efficacy scales in relation to demographic factors

Variables		Computer Experience		Computer Self-Efficacy	
		Mean	SD	Mean	SD
Mode	Paper	12.02	2.61	*24.57*	*7.89*
	Computer	12.17	2.89	24.57	*7.77*
Gender	Male	12.05	2.74	*24.83*	*7.47*
	Female	12.13	2.74	*24.29*	*8.20*
College	Sur	11.43	2.88	*26.58*	*6.42*
	Ibri	12.07	2.61	*27.00*	*5.63*
	Nizwa	12.71	2.59	*19.38*	*8.79*
Region	AlSharqiya	11.75	2.91	*25.80*	*7.17*
	AlBatinah	11.97	2.65	*23.71*	*7.95*
	AlDhahira	12.20	2.56	*26.06*	*7.13*
	AlDakhlia	13.05	2.47	*19.01*	*8.35*
Mean and SD for the whole sample		12.09	2.74	24.57	*7.83*

Table 17. *Computer Experience and Computer Self Efficacy Mean and SD, in Relation to (Mode, Gender, Collage, and Region)*

It can be seen in the table above that the computer experience mean is very similar between the paper and the computer groups as well as males and females. However, it is noticeable that Nizwa College students have a higher computer experience mean than Ibri and Sur colleges. In relation to the region factor, all three regions have similar mean except AlDakhlia which has a slightly higher mean.

Regarding computer self efficacy, the table above shows that the paper and the computer group have almost similar means, and the males' mean is similar to females'. However, it is noticeable that students at Nizwa College has a lower computer self efficacy mean than the other two colleges. In relation to the regions, it seems that AlSharqiya and AlDhahira have similar means. However, AlDakhila region has a lower mean than the others.

7.6. Comparison of Means (Computer Experience and Computer Self Efficacy)

This section investigates scale scores from different groups using an independent t-test or one way ANOVA to determine if there is a significant difference in the computer experience and computer self efficacy mean scores in relation to the variables such as mode, gender, college, and region.

7.6.1. Comparison between Paper and Computer Group

In order to test the hypotheses which is H0: *There is no difference in computer experience and computer self efficacy in relation to examinee's test mode,* preliminary analysis (Figure 26, 27) indicated computer experience and computer self efficacy are almost equal between the paper and the computer groups.

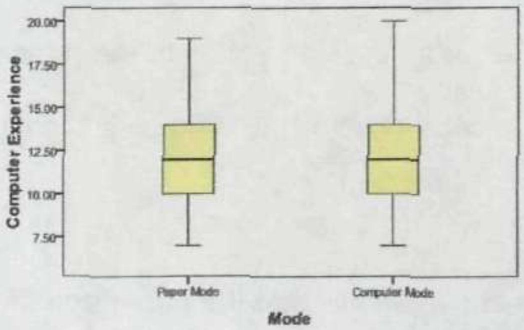


Figure 26. Mode Computer Experience Box Plots

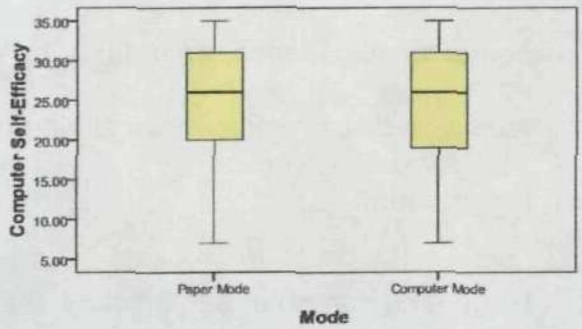


Figure 27. Mode Computer Self-Efficacy Box Plots

An independent t-test, presented in the Appendix C, showed that there were no significant differences between the means of computer experience and computer self efficacy between the paper and computer groups [computer experience *t (437) = 0.562, p= 0.574)]* and [computer self efficacy *t (437) = 0.001, p= 0.999)]* indicating the equivalence of computer experience and computer self efficacy in the paper and computer groups. It is important to know that the two experimental groups (paper or computer) did not differ in computer experience or self efficacy.

7.6.2. Comparison between Gender

In order to test the hypotheses which is H0: *There is no difference in computer experience and computer self efficacy in relation to examinee's gender,* graphical displays (Figure 28 and 29) and the summary statistics in Table 17 above suggest that computer experience and self efficacy were comparable on average between males and females.

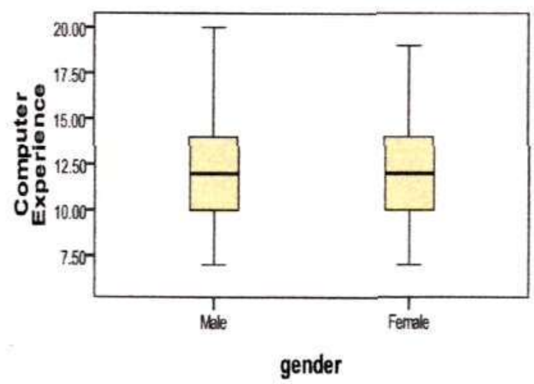


Figure 28. Gender Computer Experience Box Plots

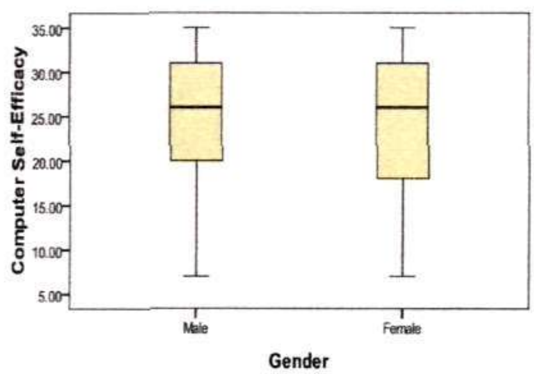


Figure 29. Gender Computer Self-Efficacy Box Plots

Independent t-tests showed no statistically significant difference [*t (439) =0.302, p=0.763)]; [*t (473) =0.722, p=0.471)] between males' and females' computer experience and computer self efficacy respectively. In addition, the effect size calculated by eta squared was very small (0.00005; 0.001). This may be surprising because I found a consistent gender difference in performance on the test and much of literature suggests that such variables as computer experience and computer self efficacy have a positive impact in the students' performance.

7.6.3. Comparison between Colleges

In order to test the hypotheses which is H0: *There is no difference in computer experience and computer self efficacy in relation to examinee's colleges*, preliminary analysis using box plots (Figure 30, 31) as well as means, standard deviations which are presented in Table 17 above suggest that Nizwa college *students have the highest computer experience and the lowest computer self* efficacy, while Sur College students have the lowest computer experience among the three colleges.

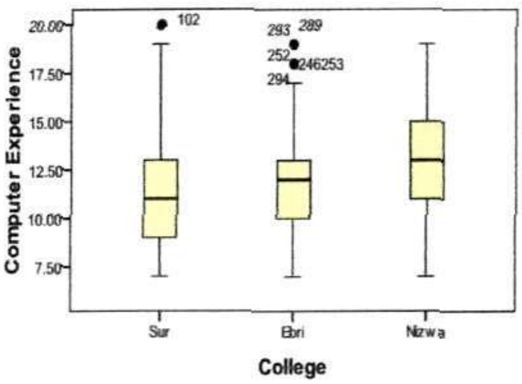


Figure 31. College Computer Experience Box Plots

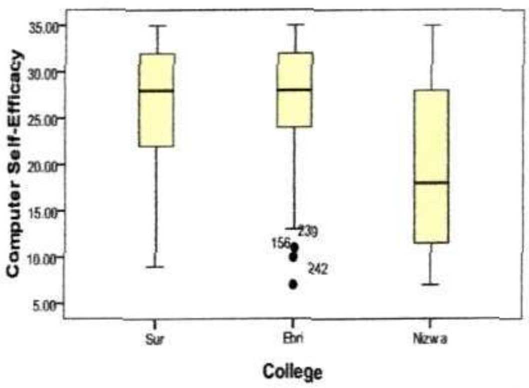


Figure 30. College Computer Self-Efficacy Box Plots

One way ANOVA shows that there was a statistically significant difference between colleges in both computer experience and computer self-efficacy (*$F_{(2,436)}$ =7.842, p< 0.0005]; (F (2,436) =53.65, p< 0.0005]* respectively. The effect size was small (Eta squared=0.04) for computer experience but it was large for computer self-efficacy (Eta squared=0.20). This is reflected in the box plots above.

Follow up analysis using the Tukey HSD test (Appendix C) indicated that Sur College students had the lowest computer experience whereas Nizwa and Ibri Colleges did not differ significantly. However, even though Nizwa college students had a higher computer experience than students at Sur college, they have the lowest computer self efficacy compared to other colleges. It is not clear, but it may be that Sur College has students from just one region (AlSharqiya), who have lower computer experience compared to the other regions, this might lead them to feel confidence and secure as most of them have low computer experience and no spirit of competition as they cannot compare themselves with students from other regions.

7.6.4. Comparison between Regions

In order to test the hypotheses which is H0: *There is no difference in computer experience and computer self efficacy in relation to examinee's region of residence,* preliminary analysis using box plots (Figure 32, 33) as well as the mean and standard deviation presented in Table 17 indicated that AlDakhlia region has the highest computer experience mean while the other three regions seem to be comparable. In contrast, AlDakhlia region has the lowest mean computer self efficacy compared to the other three regions which seems comparable.

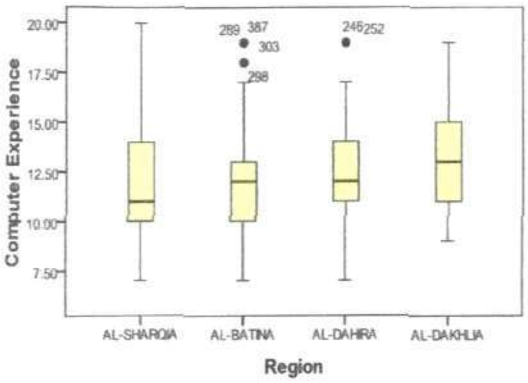


Figure 33. Region Computer Experience Box Plots

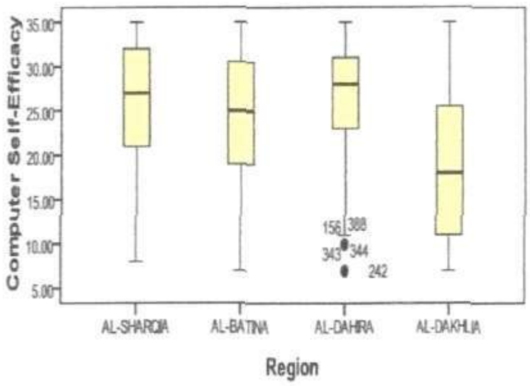


Figure 32. Region Computer Self -Efficacy Box Plots

There was a statistically significant difference between regions in the mean of computer experience and computer self efficacy *(F (3,435) =3.515, p=0.015)] ;[(F (3,435) =14.51, p<0.0005)]* respectively. However, the effect size was small (Eta squared= 0.024) for computer *experience and moderate (Eta squared=0.06) for computer self* efficacy.

Follow up analysis regarding computer experience, using the Tukey HSD test (Appendix C) indicated that there was a significant mean difference between AlDakhlia region and AlSharqiya, unsurprisingly, because most AlDakhlia region students study in Nizwa College which also had the highest mean computer experience. However, there was no significant mean difference among other regions. For computer self-efficacy, there was a significant mean difference between AlDakhlia region which had the lowest computer self efficacy mean and the other three regions.

This is not surprising, because most AlDakhlia region students are studying in Nizwa college which also had the lowest computer self efficacy mean compared to the other colleges. However, there was no significant means difference among the other three regions. It might be explained here that because students at Nizwa college have higher computer experience, so they may expected that the computerised test will require lots of complicated things and that what affected their computer self efficacy.

7.7. Comparison of Means Using Multi-Factor Analysis of Variance

In order to detect significant factors when all demographic characteristics are considered together, Multi-Factor Analysis of Variance (Table 18.19) was used to test the null hypothesis that each effect's level means are all equal, simultaneously for each of demographic variables. The assumption of normality and linearity were tested and satisfied.

Tests of Between-Subjects Effects

Dependent Variable: TOTALEXP

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Eta Squared
Corrected Model	445.036[a]	35	12.715	1.810	.004	.136
Intercept	36654.616	1	36654.616	5216.6	.000	.928
MODE	2.899	1	2.899	.413	.521	.001
GENDER	24.100	1	24.100	3.430	.065	.008
COLLEGE	106.535	2	53.267	7.581	.001	.036
REGION	60.618	3	20.206	2.876	.036	.021
MODE * GENDER	17.979	1	17.979	2.559	.110	.006
MODE * COLLEGE	10.435	2	5.218	.743	.477	.004
GENDER * COLLEGE	33.968	2	16.984	2.417	.090	.012
MODE * REGION	15.151	3	5.050	.719	.541	.005
GENDER * REGION	28.615	3	9.538	1.357	.255	.010
COLLEGE * REGION	93.037	3	31.012	4.414	.005	.032
Error	2831.675	403	7.026			
Total	67408.000	439				
Corrected Total	3276.711	438				

a. R Squared = .136 (Adjusted R Squared = .061)

Table 18. Multi Factor Analysis of Variance for Computer Experience

Tests of Between-Subjects Effects

Dependent Variable. SELFEFFI

Source	Type III Sum of Squares	df	Mean Square	F	Sig	Eta Squared
Corrected Model	8222.849[a]	35	234 939	5 079	000	306
Intercept	128857 720	1	128857 720	2785 531	000	874
MODE	4 425	1	4 425	096	757	000
GENDER	22.602	1	22.602	489	485	001
COLLEGE	2595 663	2	1297 831	28 055	000	122
REGION	371 407	3	123 802	2 676	047	020
MODE * GENDER	96 845	1	96 845	2.094	149	005
MODE * COLLEGE	20 767	2	10 384	224	799	001
GENDER * COLLEGE	635 115	2	317 558	6 865	001	033
MODE * REGION	88 599	3	29 633	638	591	005
GENDER * REGION	144 929	3	48 310	1 044	373	008
COLLEGE * REGION	180 080	3	60 027	1 298	275	010
Error	18642.640	403	46 260			
Total	291921 000	439				
Corrected Total	26865 490	438				

a *R Squared = 306 (Adjusted R Squared = 246)*

Table 19 Multi Factor Analysis of Variance for Computer Self Efficacy

When all explanatory factors are examined together there was no significant difference between the two experimental groups in terms of mode or gender There were some significant differences shown up in relation to college and region for both computer experience and computer self efficacy which confirmed the preliminary analysis using one way ANOVA The effect size was small for computer experience for college (0.036) and region (0.021). Computer self efficacy effect size was moderate for college (0.122) and small for region (0.020).

The only significant interactions which have been found in computer experience were between college and region with small effect size (0.032) The only significant interaction which has been found for computer self efficacy was between gender and college. The effect size was also small (0.033). Further exploration was done to investigate these interactions

Table 20 presents computer experience estimated marginal means and standard errors for the interaction between college and region. Sur College has students from just one region (AlSharqiya) whereas the other two colleges have students from the all four regions.

College/Region	SHRQ X(Std.E)	BATI X(Std.E)	DAH X(Std.E)	DAK X(Std.E)	Mean of College
SUR	11.43(.23)	-	-	-	11.43
IBRI	14.40(.839)	11.29(.358)	12.10(.281)	14.00(.884)	13.21
NIZWA	12.25(.469)	13.09(.462)	12.50(.501)	12.88(.375)	12.83
Mean of Region	12.95	12.27	12.50	13.61	12.09

Table 20. Estimated Marginal Means and Std.Error for Computer Experience between College/Region

The ANOVA results shows that there was a significant difference between students from AlSharqiya and AlBatinah regions who are studying at different colleges (Figure 34) whereas AlDakhlia and AlDhahira students studying at different colleges did not differ significantly. Beside, the equality of variances using Levene's test indicated that the four regions did not differ significantly in their variances (Appendix C).

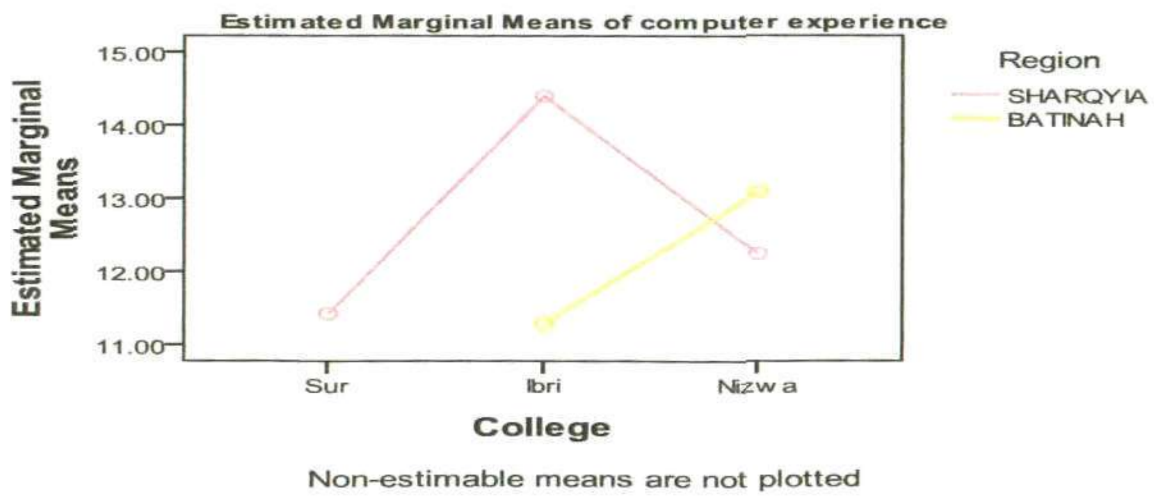


Figure 34. Computer Experience Mean plots for AlSharqiya and AlBatinah Regions' Students Studying at Different Colleges

Further investigation (Appendix C) showed that AlSharqiya students studying at Ibri College had higher computer experience than students from the same region but studying at Sur College; where AlBatinah students studying at Nizwa College had a higher computer experience than students from the same region but studying at Ibri College. An explanation can be given here that when AlSharqiya student studied at different colleges holding students from different regions that might increased the spirit of enthusiasm and competition amongst them.

Regarding the interaction between gender and college for computer self efficacy, table 21 below which displays the means for gender/college and the mean plots (Figure 35 below) indicate that females have a higher computer self efficacy mean than males in two colleges (Sur and Ibri) whereas in Nizwa college, males have a higher mean than females.

GENDER /COLLEGE	SUR X(Std.E)	IBRI X(Std.E)	NIZWA X(Std.E)	Mean of Gender
Male	25.89(0.749)	26.92(0.858)	22 01(0.772)	24.94
Female	27.78(0.980)	27.55(.690)	16.59(0.858)	23.97
Mean of College	26.83	27.23	19.30	(439)

Table 21 Estimated Marginal Means and Std.Error for Computer Self-Efficacy Based on Gender/College

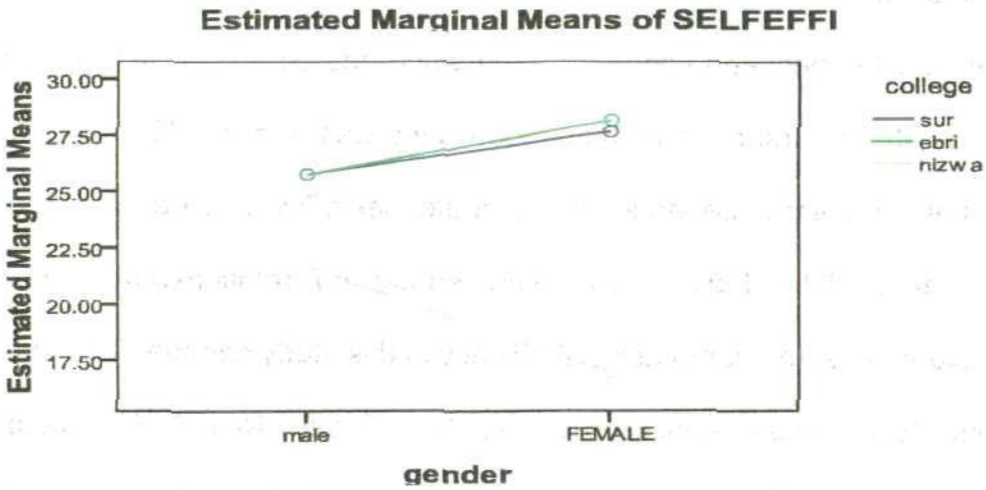
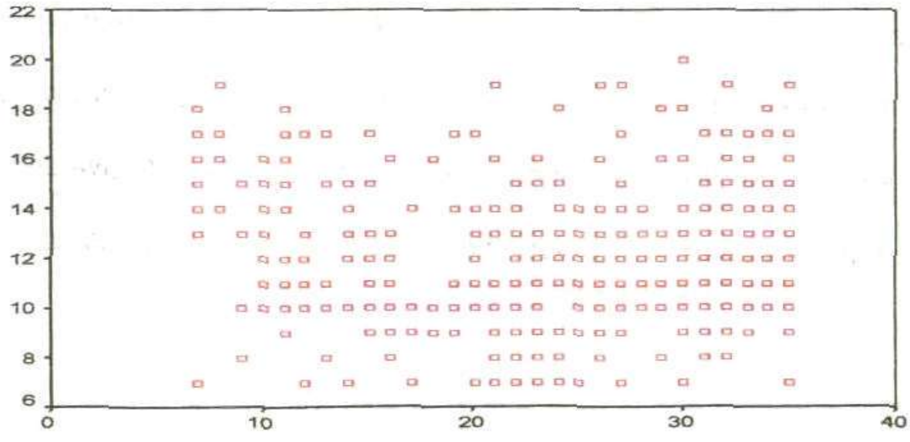


Figure 35. Gender/College Mean plots of Computer Self-Efficacy

ANOVA results show no significant difference in computer self efficacy between males and females at Ibri and Sur colleges, whereas there was a significant difference between males and females at Nizwa college were males have higher computer self efficacy than females (Appendix C). It may be explained, because females in Ibri and Sur colleges did not feel threatened from males as they have low computer experience. In contrast students at Nizwa colleges had higher computer experience which might affect females students computer self efficacy.

7.8. The Relationship between Computer Experience and Computer Self - Efficacy

One important question is whether there is a relationship between computer experience and computer self efficacy as the previous result showed that students with higher computer experience have lower computer self efficacy. However, findings in most of the literature on the topic suggest that there is a positive relationship between computer experience and computer self efficacy. In this study the Pearson correlation coefficient was $r = 0.006$ ($p=0.900$), indicating that computer experience and computer self efficacy were not related. Also the scatter plots (Figure 36) below shows no relationship between computer experience and computer self efficacy in this study sample. This study result seems to be inconsistent to many studies (Fagan, Neill & Wooldridge, 2004) that concluded the positive relationship between these two variables. However, *my study sample is different than the other studies* as most students in my study had low computer experience in general. So, some students who had low computer experience may think that it is adequate while other students who have more computer experience may think is not.



SELFEFFI

Figure 36. Scatter Graph of the Relationship between Computer Experience and Computer Self-Efficacy

7.9. The Relationship Between Computer Experience and Computer Self- Efficacy on the Total Test Performance.

Multi-Factor ANCOVA was performed to investigate whether computer experience and computer self efficacy as covariate variables have an influence on the total English test scores and the grammar and reading scores separately (Tables 22, 23 and 24). The assumptions of normality and linearity were tested and satisfied (Appendix C).

Tests of Between-Subjects Effects

Dependent Variable: TOTAL

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Eta Squared
Corrected Model	8628.291[a]	37	233.197	8.426	.000	.437
Intercept	4919.392	1	4919.392	177.749	.000	.307
SELFEFFI	185.983	1	185.983	6.720	.010	.016
EXPERIEN	167.985	1	167.985	6.070	.014	.015
GENDER	1659.908	1	1659.908	59.976	.000	.130
MODE	132.109	1	132.109	4.773	.029	.012
COLLEGE	68.332	2	34.166	1.235	.292	.006
REGION	189.648	3	63.216	2.284	.078	.017
GENDER * MODE	337.382	1	337.382	12.190	.001	.030
GENDER * COLLEGE	4.844	2	2.422	.088	.916	.000
MODE * COLLEGE	12.489	2	6.245	.226	.798	.001
GENDER * REGION	247.458	3	82.486	2.980	.031	.022
MODE * REGION	36.387	3	12.129	.438	.726	.003
COLLEGE * REGION	152.719	3	50.906	1.839	.139	.014
Error	11098.083	401	27.676			
Total	202641.000	439				
Corrected Total	19726.374	438				

a. R Squared = .437 (Adjusted R Squared = .385)

Table 22. Multi Factor Analysis for the Effect of Computer Experience and Computer Self -Efficacy on Total English Test Performance

Tests of Between-Subjects Effects

Dependent Vanable. TOTGRAMM

Source	Type III Sum of Squares	df	Mean Square	F	Sig	Eta Squared
Corrected Model	3807 138[a]	37	102 896	6 264	000	366
Intercept	1820 720	1	1820 720	110 841	000	217
SELFEFFI	67 348	1	67 348	4 100	044	010
EXPERIEN	82.746	1	82.746	5 037	025	012
GENDER	726 618	1	726 618	44 235	000	099
MODE	41 259	1	41 259	2.512	114	006
COLLEGE	12.262	2	6 131	373	689	002
REGION	91 064	3	30 355	1 848	.138	014
GENDER * MODE	326 977	1	326 977	19 906	000	047
GENDER * COLLEGE	685	2	342	021	979	000
MODE * COLLEGE	1 371	2	686	042	.959	000
GENDER * REGION	111 580	3	37 193	2.264	081	017
MODE * REGION	26 325	3	8 775	534	659	004
COLLEGE * REGION	80 561	3	26 854	1 635	181	012
Error	6586 985	401	16 426			
Total	82059 000	439				
Corrected Total	10394 123	438				

a. R Squared = 366 (Adjusted R Squared = 308)

Table 23. Multi Factor Analysis for the Effect of Computer Experience and Computer Self -Efficacy on Grammar

Tests of Between-Subjects Effects

Dependent Vanable TOTREAD

Source	Type III Sum of Squares	df	Mean Square	F	Sig	Eta Squared
Corrected Model	1209 032[a]	37	32.677	4 845	000	309
Intercept	759 251	1	759 251	112 569	000	219
SELFEFFI	30 153	1	30 153	4 471	035	011
EXPERIEN	14 463	1	14 463	2.144	144	005
GENDER	190 941	1	190 941	28 310	000	066
MODE	25 639	1	25 639	3 801	052	009
COLLEGE	26 895	2	13 448	1 994	138	010
REGION	18 237	3	6 079	901	441	007
GENDER * MODE	9 282E-02	1	9 282E-02	014	907	000
GENDER * COLLEGE	8 917	2	4 458	661	517	003
MODE * COLLEGE	7 033	2	3 516	521	594	003
GENDER * REGION	32 978	3	10 993	1 630	182	012
MODE * REGION	1 585	3	528	078	972	001
COLLEGE * REGION	23 703	3	7 901	1 171	320	009
Error	2704 653	401	6 745			
Total	29508 000	439				
Corrected Total	3913 686	438				

a R Squared = 309 (Adjusted R Squared = 245)

Table 24. Multi Factor Analysis for the Effect of Computer Experience and Computer Self -Efficacy on Reading

There were significant effects of computer experience and computer self efficacy on the total English test scores *[F=6.070; p=0.014; F=6.720, p=0.010]* and the grammar *test [F=5.037; p=0.025; F=4.100, p=0.044]*. However, in the reading test, there was no significant effect for computer experience *[F=2.144,* *p=0.144]* whereas there was a significant effect of computer self efficacy [F=4.471, p=0.035].

It is worth noting that when computer experience and computer self efficacy were included as covariate variables, no college and region significant difference in the total English test score was detected. However, the mode difference, gender difference and differential mode effect of the two genders were still present.

The parameter estimates (Table 25) showed a positive relationship between computer experience and the total English test performance and the grammar scores. However, there was a negative relationship between computer self efficacy and the total test performance, grammar and reading scores. That means students who report low computer experience and high computer self efficacy performed lower on the English test.

Parameter Estimate(sig)	Total performance	Grammar	Reading
Computer Experience	0.25(0.01)	0.17(0.02)	0.072(0.14)
Computer Self Efficacy	-0.10(0.01)	-0.06(0.04)	-0.04(0.03)

Table 25. Parameter Estimate between Computer Experience and Computer Self-Efficacy and Total Test, Grammar and Reading

7.10. Discussion

The findings presented in this chapter may explain some differences relating to test performance which were found in the previous chapter such as college and region differences. As the previous chapter shows, there was a significant college difference with Nizwa College achieving better scores compared to Sur and Ibri Colleges. In this chapter it was found that students at Nizwa College have higher computer experience than those in the other two colleges. Also this chapter found that there is a positive significant relationship between computer experience and the total test performance. All of these may explain why Nizwa College students performed better in the test, as they had higher computer experience compared to the others

In addition, in the previous chapter it was found that AlDakhlia students' performance was higher on the total test compared to AlSharqiya region, which may be explained by the higher computer experience level of students from AlDakhlia region. Also, when computer experience and computer self efficacy are included as covariate variables, no college and region differences in the total test score have been detected. This suggests that the performance differences between students at different colleges, or from different regions, may be explained by their computer experience, rather than any feature of the college or region.

It was surprising that Nizwa college students had the lowest computer self efficacy level compared to the other two colleges even though they had the highest computer experience level. Also, the present study found no relationship between computer experience and computer self efficacy in contrast to many studies that did find a relationship (Busch, 1995; Johnson, Ferguson & Lester, 1999; Moroz & Nash, 1997). Even though Fagan, Neill & Wooldridge (2004) point out that increased familiarity with computers is positively related to computer self efficacy, there are other studies which concluded that the increased computer experience does not necessarily translate into increased computer self-efficacy (Sam, Othman & Nordin, 2005).

However we may be able to relate our findings to the lack of student training in using computerized assessment, which may affect student confidence. Moreover, computer self efficacy depends on self beliefs and perceptions more than knowledge or experience. Also, students who think they are capable of doing tasks will be likely to do so successfully. In addition, it might be explained by motivation as Bandura (1977) points out that "self efficacy is believed to play a critical role in self-motivation, especially when a certain level of motivation is necessary to initiate coping with unfamiliar tasks. As Sein, Olfman & Bostrom (1987) suggested that focusing on the development of an appropriate motivational level is more important than knowledge acquisition "(Cited in Torkzadeh, Koufteros & Pflughoeft, 2003, p.263, 264).

The other finding relates to the impact of computer self-efficacy on the students performance which had a small but significant negative impact on the students' total scores. This finding is actually inconsistent with the findings achieved by most of the studies that have investigated the impact of computer self-efficacy on students' performance (Chou, 2000; Fagan, Neill & Wooldridge, 2004; Jonhson, Ferguson & Lester, 1999).

In fact, most of the studies cited above have concluded that increased computer self-efficacy levels had substantial effects on reducing computer anxiety and, accordingly, improving student performance on computer-related tasks (Khorrami-Arani, 2001). This finding is stressed by Torkzadeh, Koufteros & Pflughoeft who assert that "researchers to date confirm that computer self-efficacy not only determines decisions by individuals to accept and use the computer system, but is also a good predictor of achievement in computer-related tasks" (Torkzadeh, Koufteros & Plughoeft 2003, p.264)

However the result in my study might be explained by students' inexperience in the use of computers as assessment tools, which is due to lack of computer experience in general

It should also be stressed that although lack of experience with computers tends to affect student performance on computerised assessment, there is still a wide consensus on its negligible impact on the performance of students taking multiple-choice questions Citing (Bennett, 2002; Bridgeman, Bejar & Friedman, 1999, Taylor *et al.*, 1998), Paek points out that "for multiple-choice tests, the research suggests that differences in computer experience have virtually no effect on test scores" (2005, p.10) However the present study found a weak but significant relationship between computer experience and test performance. This agrees with the findings of numerous studies like Choi & Tinkler (2002) who point out that computer test performance was impacted by computer experience, and that the students who had a lower computer experience level tended to attain lower scores than those having a higher computer experience level when taking computerised tests According to Peak (2005), "Choi & Tinkler found that computer experience was related to computer test performance, as students who rarely used a computer tended to perform lower in both mathematics and reading than those students who had more computer experience" (2005, p 15).

However, Choi & Tinkler note that acquiring experience with computers would ultimately translate into students' attainment of better scores on CBTs Hence, according to them increased student computer experience would account for eliminating the 'testing mode effect'. Choi & Tinkler point out that "more frequent exposure to such online exams may eventually eliminate the novelty and mode effect" (2002, p.10).

Further investigation showed that students from AlSharqiya and AlBatinah regions studying at different colleges have different computer experience levels; whereas students from AlDakhlia or AlDhahira who are studying at different colleges have the same computer experience level. As Ibri and Nizwa colleges have students from four regions that might lead to increase the motivation and competition between students contrary to Sur College which have students just from Alsharqiya region.

Also this study found a significant interaction in computer self efficacy between gender and college. Further investigation showed that while in Nizwa College male students had a higher computer self-efficacy level than females, there was no significant difference between males and females at Ibri and Sur colleges. It should be noted that although this result generally agrees with the findings of most of the studies which have addressed the issue of computer self-efficacy in relation to gender, part of this result is nonetheless totally inconsistent with much of the literature written on this issue. That is, while the first part of the result (the case of Nizwa College) seems to be in harmony with the findings postulated in most previous studies which have investigated the impact of gender on attaining computer self-efficacy and have stressed that male, rather than female, students were the ones had a higher computer self-efficacy level (Betz & Hackett, 1981; Busch, 1995; Comber *et al.*, 1997; Hackett, 1985; Isiksal & Askar, 2005; Post-Kammer & Smith, 1985) ; there is little empirical evidence to support the equivalence of computer self efficacy between males and females in both Ibri and Sur Colleges. This, in fact, contradicts Busch (1995) who reports tracing no gender differences in computer self efficacy regarding simple computer tasks. However, the finding of my study is to a great extent consistent with the finding of Comber *et al.* (1997) whose study traced some gender difference in favour of males who tended to have more confidence and experience in using computers despite the fact that they did not find any significant gender differences when they used long experience as a covariate.

Many studies have also reported that males have more computer experience and more access to the computers than females (Clariana & Wallace, 2002; Comber *et al.*, 1997, Wallace & Clariana, 2005; Wolfe & Manalo, 2004) and pointed out that this might affect female's performance in computerized tests. However, the present study did not find any differences between males and females regarding computer experience or computer self efficacy (except Nizwa College), which did not explain why females did better in the paper test than in the computerised one while males achieved better scores in the computer rather than the paper test.

The question that now arises is whether the gender difference in performance on paper or on computer is related to actual computer experience, or whether there are more deep-seated attitudes towards computers, especially when used for assessment, which may affect performance. This is what the next chapter intends to explore.

Chapter Eight: Students' Attitude towards Computer Assisted Assessment (CAA)

8.1. Introduction

Previous parts of this study have found college and region test performance differences which were associated with differences in students' computer experience. Also there were mode effects on test performance, which has indicated that the higher computer experience is, the better test performance gets. However, results did not explain why males did better on the computer while females did better on the paper test, even though they generally have comparable computer experience and computer self efficacy. Therefore, the main purpose of this chapter is to explore whether males and females are different in their attitudes towards computerised assessment. Also, this chapter will seek to explore whether such variables as college or region affect students' attitude and perceptions of Computer Assisted Assessment (CAA).

The following chapters of this study will be dealing with both quantitative and qualitative analysis methods to explore more deeply students' perceptions and views towards computerised tests. Quantitative data obtained from the questionnaire may be explained and clarified more through the focus groups. The focus groups will give us deeper insight into students' attitudes towards computerised tests and may present us with clarifications which cannot be obtained from quantitative data.

As done previously, the same statistical procedures were applied throughout this chapter. In addition, qualitative approaches were used to investigate the research questions about the students' perceptions, thoughts, views and feeling towards implementing CAA Finally, the findings will be discussed in the light of the literature.

8.2. Sample Characteristics

The sample sizes are different from previous chapters because only students who sat the computerised test participated in the questionnaire and focus group interviews. The table 26 below shows the sample sizes and percentages for students' that participated to complete the Attitudes Towards Computerised Assessment Scale (ATCAS) in relation to gender, college and region.

Variables		Count	Percentage
Gender	Male	95	48.5%
	Female	101	51.1%
College	Sur	62	31.6%
	Ibri	73	37 2%
	Nizwa	61	31.1%
Region	AlSharqiya	81	41.3%
	AlBatinah	31	15.8%
	AlDhahira	54	27.6%
	AlDakhlia	30	15.3%
Total students sample		196	100%

Table 26. Sample Size and Percentages According to Gender, Collage, and Region

Of the 196 students who filled out the questionnaire, 95 (48.5%) were males and 101 (51.5%) were females. This cohort comprises students belonging to three Applied Science Colleges, 62 students from Sur, 73 from Ibri and 61 from Nizwa which respectively make up 31.6%, 37.2% and 31.1% of the total student number. Once again we should note that these students come from different regions of the Sultanate as 81 students (41.3%) come from AlSharqiya, 31 students (15.8%) come from AlBatinah, 54 students (27.6%) from AlDhahria and 30 (15.3%) come from AlDakhlia region.

8.3. Factor Analysis and Reliability of the Attitude Towards Computerised Assessment Scale (ATCAS).

The ATCAS scale contains 13 items developed by Smith and Caputi (2004). All 13 items of the ATCAS were subjected to principal component analysis. The Kaiser-Meyer-Oklin value was 0.75 exceeding the recommended value 0.60 (Kaiser, 1974) and the Barlett's test of sphericity reached statistical significance supporting the factorability of the correlation matrix.

Principal components analysis revealed the presence of five components with eigenvalues exceeding 1 explaining 27.1 percent, 11.1 percent, 10 percent, 8.7 percent and 7.8 percent of the variance respectively. The total variance explained by these components was 64.7 percent (Appendix D).

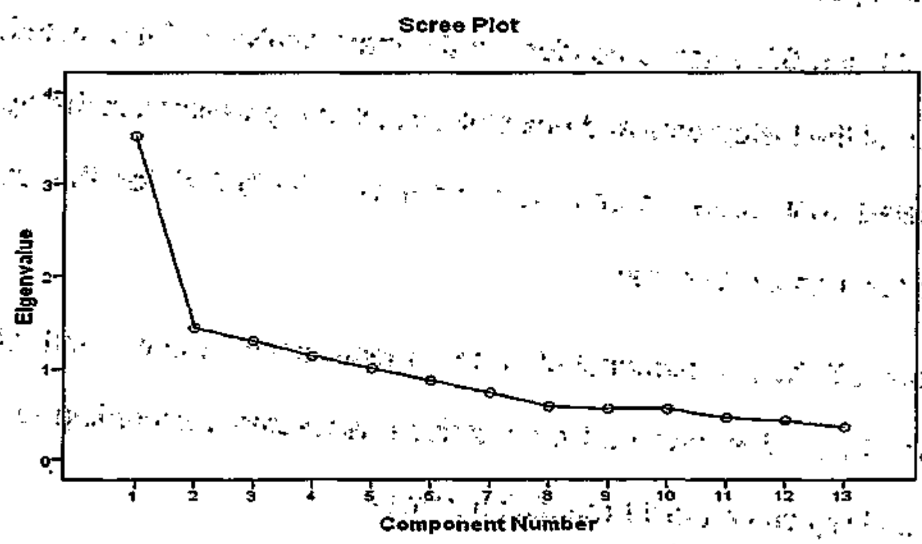


Figure 37. Scree Plots of Attitudes Towards Computerised Assessment Scale (ATCAS)

However, the inspection of the scree plot (Figure 37) reveals a clear break after the first component. It was decided to retain one component for further investigation. The one component solution explained a total of 27.1 percent of the variance (See Appendix D). Most items loaded on this component with a coefficient of more than 0.40 except two items (5 and 8) which had low coefficient (0.25 and 0.16). The item number 5 had a low coefficient in the pilot study and this larger sample confirmed the finding. The internal consistency of the scale (Cronbach's alpha) increased from 0.76 to 0.78 when these two items were deleted. Also, when these two items was deleted, the one component solution explained 31.52 percent of the variance. Because of the low loading and the increase in Cronbach's alpha, 11 out the 13 items on the ATCAS questionnaire were retained for further analysis.

8.4. Preliminary Analysis and Descriptive Statistics for Students' Attitudes Towards Computer Assessment Scale (ATCAS)

The scale items were coded in order to let higher score reflect positive attitudes towards computerised test. The individual marks were summed to give score between 11 and 55. Prior to conducting the main analysis of the attitude data, inspection of the histograms and box plots, which are presented in the Appendix D, indicated that scale scores were normally distributed for the students' attitudes towards CAA scores.

The table 27 below demonstrates the mean scale scores and standard deviation of attitudes towards Computerised assessment according to gender, college and region on the 11 remaining items.

Variables		Mean	SD
Gender	Male	32.46	7.04
	Female	29.31	9.19
College	Sur	29.74	8.18
	Ibri	30.16	9.32
	Nizwa	32.75	6.99
Region	Alsharqiya	30.36	8.10
	AlBatinah	31.09	5.918
	Aldhahira	30.83	10.22
	AlDakhlia	31.87	7.64
Mean and SD for the whole students sample		30.84	8.35

Table 27. Students' Scores on ATCAS Questionnaire (Mean, SD) According to Gender, College, and Region.

Preliminary analysis using box plots, which are presented in Appendix D, and the means and standard deviations presented in the table 27 above, indicated that males reveal slightly more positive attitudes than females. Females also showed slightly greater variance. As for the college variable, it could be noticed that while there is almost a similar mean for students' attitude towards CAA at Ibri and Sur colleges, the attitude of the students at Nizwa College was slightly higher. In regard to the regions variable, it seems obvious that there was almost no difference between the students from different regions.

8.5. Comparison of Means (ANOVA and Multi Factor Analysis of Variance)

This section will compare the mean scores of different groups to determine if there are any significant differences between the mean scores in relation to variables such as gender, college, and region. As done in the previous chapters, the means of each factor will be compared one by one by t-test and ANOVA, then all together by Multi Factor Analysis Of Variance. It important to note that the statistical analysis based on computer group sample only, so mode effect cannot be included

In order to test the hypothesis which is *Ho: There is no difference between students' attitudes towards Computer Assisted Assessment in regard to the gender, college or region of origin*, for gender variable, t-test found a significant difference [t (186) =2.71, p=0.007] indicating that males' attitude was more positive (X= 32.46, SD =7.39) than females (X= 29.31; SD=9.19). Since the assumption of equal variances was violated in the total test score [Levene's test F=5.673, p=0.018], the result presented is based on an alternative t-value which compensates for the fact that the variances are not the same (Appendix D). However, the magnitude of the difference was small (Eta squared=0.036) according to Cohen's (1988) criterion.

One way ANOVA, presented in the Appendix D, found no significant difference in the students' attitudes toward the computerised test between colleges *[F (2,193) = 2.413, p = 0.092]*, and the actual difference in mean scores between the groups was small (Eta squared= 0.024) according to Cohen's (1988) terms.

A further ANOVA, presented in the Appendix D, found no significant difference between regions *[(F (3,192) =0.248, p=0.863)]*, and the actual difference in the mean scores between the groups was extremely small (Eta squared=0.004) according to Cohen's (1988) terms.

Multi-Factor Analysis of Variance was used (Table 28) to test the null hypothesis that each effect's level means are all equal, when all of the demographic variables are considered together. The assumption of normality and linearity was tested proved satisfactory.

Tests of Between-Subjects Effects

Dependent Variable :ATCAS

Source	Type III Sum of Squares	Df	Mean Square	F	Sig.	Eta squared
Corrected Model	1511.060[a]	17	88.886	1.310	.191	.111
Intercept	120366.627	1	120366.627	1773.656	.000	.909
Gender	10.420	1	10.420	.154	.696	.001
College	181.304	2	90.652	1.336	.266	.015
Region	92.799	3	30.933	.456	.714	.008
gender * college	.020	2	.010	.000	1.000	.000
gender * region	208.024	3	69.341	1.022	.384	.017
college * region	317.465	3	105.822	1.559	.201	.026
Error	12079.716	178	67.864			
Total	199968.000	196				
Corrected Total	13590.776	195				

a. R Squared = .111 (Adjusted R Squared = .026)

Table 28. Multi Factor Analysis in Attitudes Towards Computerised Assessment scale (ATCAS)

When all explanatory factors are examined together there was no significant effect of any of the factors as well as no significant interaction. It should be noted this result is different from the conclusion based on just a simple t-test between gender and consistent with the college and region result based on one way ANOVA. So, the small gender difference can be explained by different samples across regions and colleges, rather than a student gender difference. Students' differences are therefore considered across gender, region and college.

8.6. The Relationship between Students' Attitude towards Computerised Testing and Their Computer Experience and Computer Self-Efficacy

Students' attitudes towards computerised testing may be affected by their previous computer experience or self-perception of computer skills. Therefore, I explored the relationship between attitudes towards computerised testing and computer experience and computer self efficacy using the Pearson Correlation Coefficient (Table 29).

Variables		Computer Experience	Computer Self –Efficacy
ATCAS	R	0 099	-0 063
	SIG	0 167	0 384

Table 29. Correlation Coefficient Between ATCAS and Computer Experience, and Computer Self Efficacy

The table 29 above shows that students' attitudes towards computerised testing were not related to computer experience and computer self efficacy which is inconsistent with many studies that concluded a positive relationship between students attitude towards computer and student computer experience or computer self efficacy.

8.7. Effect of the Students' Attitude Towards Computerised Test on the Test performance

Even though there does not seem to be any difference in attitudes between genders across colleges and regions, it may still be that attitude does affect performance as students with more positive attitudes may have scored better.

As a preliminary analysis, Pearson correlation coefficients were derived to explore the relationship between performance on the English test, grammar and reading and students' attitudes towards computerised testing. Table 30 shows that students general attitude towards computer testing was not associated with their performance on the total test, grammar or reading.

Variables		Total English Test	Grammar	Reading
ATCAS	R	*0.005*	0.000	0.012
	SIG	*0.914*	1.000	0.868

Table 30. Correlation Coefficient Between ATCAS and the Total Test, Grammar and Reading

Then, Multi Factor ANCOVA (Table 31) was used to investigate the relationship between students' attitudes towards computer testing as a covariate variable and their scores on the total English test, grammar and reading. The assumption of normality, linearity and homogeneity of variances were tested and satisfied.

Tests of Between-Subjects Effects

Dependent Variable Total English Test Score

Source	Type III Sum of Squares	Df	Mean Square	F	Sig	Partial Eta Squared
Corrected Model	2002 717[a]	20	100 136	2 909	000	.249
Intercept	1086 407	1	1086.407	31.558	.000	.153
C Self efficacy	7 332	1	7 332	213	645	.001
C.Experience	19 809	1	19 809	575	449	.003
C Attitude	41 646	1	41 646	1 210	273	.007
Gender	263 689	1	263 689	7 660	006	042
College	64 126	2	32 063	931	396	011
Region	74 152	3	24 717	718	542	012
gender * college	44 193	2	22 096	642	528	007
gender * region	110 372	3	36 791	1 069	364	018
college * region	304 743	3	101 581	2 951	034	048
Error	6024.487	175	34 426			
Total	88278 000	196				
Corrected Total	8027 204	195				

a R Squared = 249 (Adjusted R Squared = 164)

Table 31 Multi-Factor ANCOVA Between ATCAS on Total English Test Score

There was no significant relationship between students' attitudes toward computerised testing and their performance on the total English test, grammar and reading.

However it worth noting that when attitudes towards computerised testing is dealt with as a covariate variable, college and region showed no significant difference in the total English test, grammar and reading scores. However, there was a gender difference in the total score and the reading score while no gender difference has been detected in the grammar test. (See Appendix D, for grammar and reading Multi Factor ANCOVA).

Also there was a region/college significant interaction in the total test score and the reading. Further exploration using ANOVA test, presented in Appendix D, revealed that there was a significant difference between the students from AlSharqiya and AlBatinah regions who are studying at different colleges whereas AlDakhlia and AlDhahira students studying at different colleges did not differ significantly.

Follow up analysis using Dunnett's C (equal variances not assumed) revealed that AlSharqiya students studying at Nizwa College had higher scores than the students from the same region who are studying at Sur College.

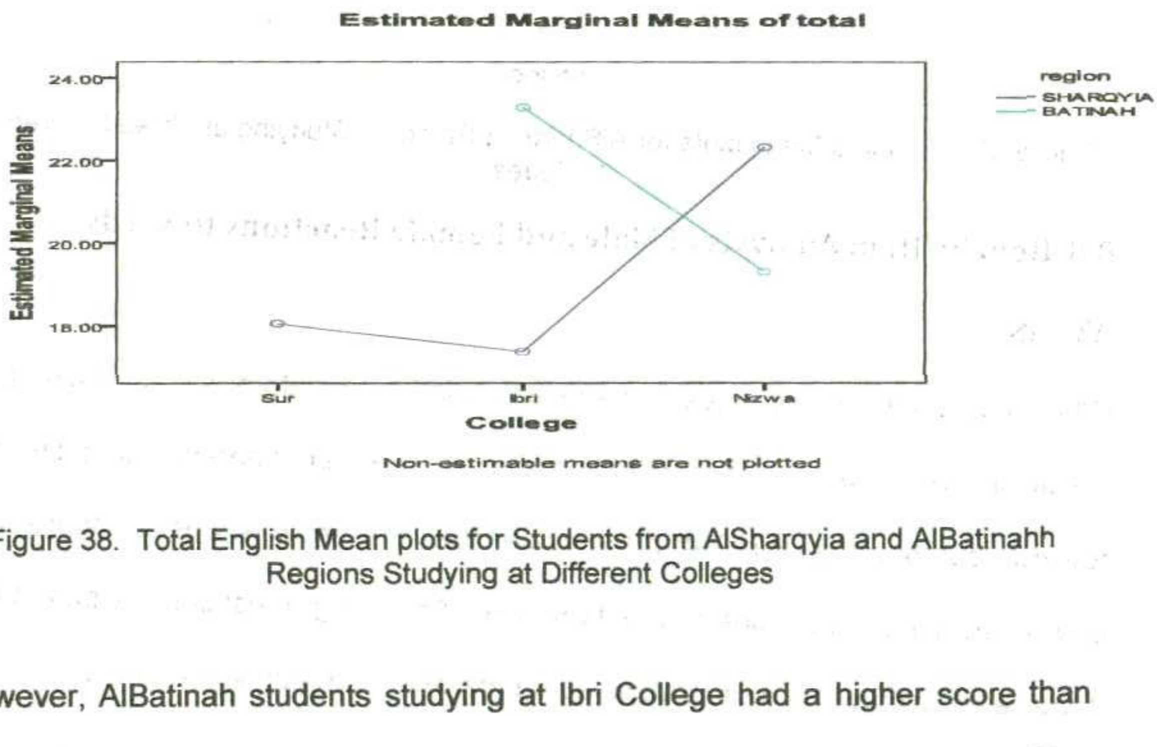


Figure 38. Total English Mean plots for Students from AlSharqyia and AlBatinahh Regions Studying at Different Colleges

However, AlBatinah students studying at Ibri College had a higher score than the students from the same region who are studying at Nizwa College (See Appendix D). It should be noted that no students from AlBatinah are studying at Sur College (Figure 38).

In the reading test, ANOVA test, presented in Appendix D, showed that there were significant differences between students from AlSharqiya region studying at different colleges (Figure 39) while there were no differences between the students from the other three regions who are studying at different colleges. Follow up analysis using Tukey HSD revealed that AlSharqiya students who are studying at Nizwa College had higher score compared to the students from the same region but studying at Sur College (See Appendix D).

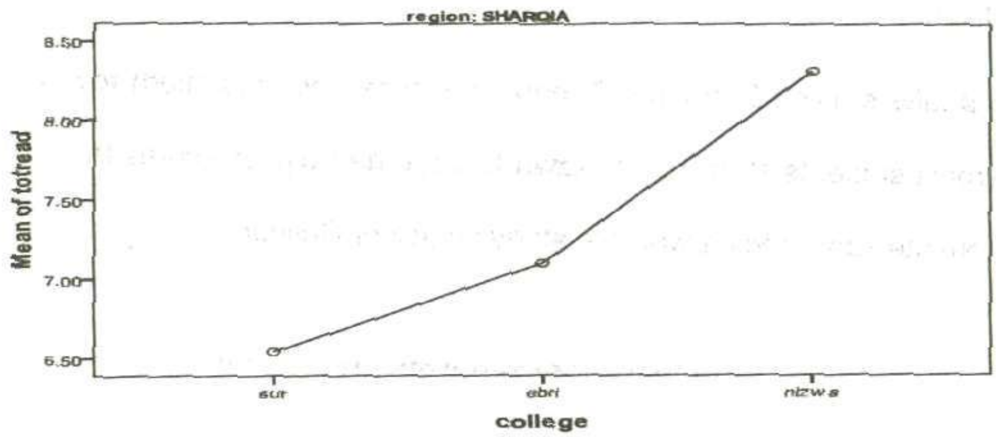


Figure 39. Reading Mean plots for AlSharqiya Students Studying at Three Different Colleges

8.8. Item by Item Analysis of Male and Female Reactions towards ATCAS

When the items on the questionnaire are combined onto a single scale, the result shows weak evidence (Section 8.5.) of gender difference in attitudes towards computer testing. What the scale does not show is whether there are any differences in attitudes at the item level. Analysing questionnaire based on items level will give us more details about males and females responses.

The result in the table 32 below presents the number of participants (males and females) selecting agree (i.e., 'agree' or 'strongly agree') or disagree (i.e., 'strongly disagree' or 'disagree'). The first glance at males' and females' responses reveals that females feel more nervous and anxious than males. In addition, a greater proportion of females found it easier to check responses on the paper rather than on the computer test and felt more comfortable completing the test on paper than on the computer. Also, while 57% of the female participants found that reading questions from the computer screen more difficult than reading questions from the paper test, only 32% of the male participants reported the same concern. However, the preference for taking computer test in the future was almost similar between males and females.

Items		Agree n(%)	Disagree n(%)	Chi Square p-value
(1) More nervous on computer than paper	Male	18 (18.9)	53(55.8)	.006
	Female	36 (35.6)	38(37.6)	
(2) Test instructions on the computer were difficult to understand	Male	31(32.6)	38(40)	.574
	Female	30(29.7)	47(46.5)	
(3) Helpful if more practice time was given before starting the test	Male	62(65.3)	20(21.1)	.518
	Female	68(67.3)	16(15.8)	
(4) More difficult reading question on the computer than paper	Male	31 (32.6)	42 (44.2)	.004
	Female	58 (57.4)	29 (28.8)	
(6) More anxious on computer than paper	Male	22 (23.2)	49 (51.6)	.270
	Female	32 (32)	46 (46)	
(7) Lack of computer experience interfered with performance on computer test	Male	24(25.3)	45(47.4)	.772
	Female	30(29.7)	48(47.5)	
(9) Computerised test require too much computer experience	Male	42(44.2)	30(31.9)	.202
	Female	59(58.4)	36(25.7)	
(10) Wish computerised test did not bother me so much	Male	49(52.1)	25(26.6)	.004
	Female	73(73)	11(11)	
(11) Worrying about pressing the wrong key	Male	37(38.9)	38(40)	1.000
	Female	40(39.6)	40(39.6)	
(12) Easier to check my responses on P&P test rather than computer	Male	41 (43.2)	34 (35.8)	.219
	Female	60 (59.4)	32 (31.7)	
(13) More comfortable completing the test on paper than computer	Male	44 (46.3)	29 (30.5)	.068
	Female	61 (60.4)	19 (19.8)	

Table 32. Proportion of Males and Females Selecting Agree and Disagree Response on the ATCAS

Using Chi Square Tests to investigate whether females are more likely than males to have negative feelings towards computerised assessment showed that the proportion of females who were more nervous and found it difficult to read questions from the computer screen, and also wished computerised test did not bother them scored significantly higher than the proportion of males. However, the proportion of females that were more anxious and uncomfortable completing the test on the computer or preferred to take the paper test rather than the computer one, or would prefer not to take the computerised test was not significantly different from the proportion of males. These differences, in only three of the 11 items in the scale, may explain why an overall difference between both genders was not established when the items were combined.

Generally, as more females were nervous taking the computer test, found it more difficult to read questions from the computer screen and wished that the computerised test had not bothered them so much, it seems now evident to somewhat why females did better in the paper rather than the computer test while males did better in the computer rather than the paper one. The next section will aim to gain a deeper insight into students' attitudes towards computerised testing using focus group discussions and to check the result consistency as well.

8.9. Pre and Post -Test Focus Groups Discussion

I conducted three pairs of focus group discussions, one pair in each college. One discussion was held before the computerised test and one after. Every focus group included 10 to 12 students (males and females). Only the students who had taken the computer test participated in the focus groups as these students experienced the computer test and so they were more capable than the others of describing their feelings towards that experience. The same students at each college participated in the focus group discussions before sitting for the computer test and after completing the test. The aim behind doing so was two-fold: first to explore more deeply their attitudes towards taking a computerised test, and second to detect any change of attitude after the taking the computerised test.

When asked about their computer experience, there were no differences traced between males and females as most of them described their computer experience as moderate to good, except two males who described their computer experience as high. However, it was noticed that females were less confident dealing with the computer than males. It also was obvious that students from Nizwa and Ibri colleges had higher computer experience levels than students from Sur College and that was deduced from their answers. However, as all students were not familiar with computerised assessment, therefore, it seems as if their confidence was affected by the novelty of the mode of administration.

The focus group discussions were recorded, transcribed and then analyzed thematically. Six key themes emerged from the analysis: *gender differences, CAA familiarity, grammar and reading comprehension, language effect, system reliability, and CAA strength and weaknesses.*

8.9.1. Gender Difference

In the pre-test focus group discussions most students were worrying about this new kind of assessment. However it was quite noticeable that female students were much more worried and anxious about taking the computerised test. In contrast; although males also had some worries they were more enthusiastic and motivated to undertake this new experience. In the post-test focus groups it was obvious that males' worries and anxiety had largely vanished, whereas females still had the same feeling of worry and anxiety. They still expressed less preference for taking computerised tests as they mentioned that they are very anxious dealing with computer tasks. One female asserted that *"I have adequate computer experience but I did not feel comfortable completing the test by computer"* and all females agreed. Another female said "even *if I did well in computerised test, but I still preferred paper and pencil test"*.

However, males were less anxious and nervous than females and their computer test preference was more than that of females. One of the males stated *"if I had good training in the computerised test programme I would choose to take all my tests by computer"*. Another male said *"if I take many tests through computer that will make CAA easier for me than paper"*. One more also said *"I was not afraid or worried either before or after the test at all"*. This student previously described his computer experience as high.

Before sitting for the computerised test students' (both males and females) worries revolved around a number of points in regard to computerised testing:

- Difficulty to read from the screen, especially long texts.
- Time pressure especially for those who are slow in using or typing on a computer.
- Nature of the questions they justified their worries with comments such as *"we are used to paper-pencil test, so for the computerised test the*

nature of its questions might affect our use of the computer and our performance".

- Novelty of the method as one of the students said "*It is a new method and we are not used to it*"

- Technical problems with computers.

After the computer test, males were more confident than before while females were not feeling comfortable yet and still had worries. They justified their worries by the difficulty to read from the screen and its effect on sight due to the need to concentrate on what is written as well as back pain due to sitting for a long time in front of the screen.

8.9.2. CAA Familiarity

It is important to remember that CAA is a novel experience for these students as none of them had experienced it before. Besides, they did not undergo any practice before taking the computerised test. However, they did have a session explaining how to answer questions using the computer. Most students were anxious before the test and said they were used to paper-and-pencil tests since first grade so it would be difficult for them to take a test using a new method.

When asked in the pre-test focus group about which mode they would choose in the future if they are given the chance to choose, unsurprisingly nearly 90% of the students (31 out of 34) said they would choose the paper-and-pencil test. Their reasons were: the complexity of using the computer; the fact that computers sometimes do not work well; worrying about pressing the wrong key or button; and their familiarity with the paper pencil test. Hence, it seemed that lack of computer familiarity was the main reason many of them expressed for choosing the paper test, and because they were used to it. For example, one student said *"I prefer the paper-pencil test because I am not good at typing using the keyboard"*, and another one said *"I prefer paper-pencil test because I am slow in using the computer which will waste a lot of time and as a result I will not be able to finish it on time"*

As for the few students who chose the computerised test, one of them said *"It is very accurate in marking and giving the results and it will never be biased"*. Another one added *"I will get the results fast and that is an advantage where I do not have to wait for weeks to get them as in the paper-pencil test"*. They summarised by mentioning that it is easy, accurate and can be trusted in terms of marking During the discussion, these students described their computer experience as good or high.

It was interesting that in the post-test focus groups 56% of students (19 out of 34) would now have chosen the computerised test. Those who chose to be tested by the computer preferred that because it is no longer a worrying experience. One of the students said *"I have become more confident after doing the computerised test and my confidence has increased a lot on this type of test"*. Many students agreed that their confidence had increased after doing the test, while before doing it, they had a mixture of feelings like: worry, anxiety and fear towards the type of questions, and their nature and quality, and how to use the computer. After doing the computerised test most of those feelings disappeared and confidence replaced them. Students also confirmed that their confidence would increase more if they frequently do this kind of test until it becomes easy and normal (familiarity). As one of students said *"if I continue to* *take my test through computer frequently then computerised tests will be as easy for me as paper test."*

Also, one student said *"I was afraid before the test, but after trying it was fun and normal".* After doing the test, the majority of the students confirmed that there was nothing to be afraid of, they were comfortable, and that the test was a fun and an exciting experience. One student concluded by saying *"If I were provided with the right training I would have preferred to use the computerised test".* In different words I will accept the CAA if I was familiar with it. This result was also confirmed by the questionnaire responses as well. 69% of the students responded that they would have found it helpful if they had been given more practice in the computer before starting the test. (See Appendix D)

As for those who chose the paper-pencil test the in the post test focus group, the majority of them were females. They preferred paper and pencil test as it is the method that they were used to since they were young. One of these students said *"I used to do my test by paper since grade one so I am comfortable using it". Another added "Since I started school the only way of assessing me was through paper test so what the point of change it as I am familiar with".*

8.9.3. Grammar and Reading Comprehension

Before students sat for the test I asked them what they thought would be easier in the computerised test, grammar or reading. 70% of the students (24 out of 34) assumed that grammar would be easier. They justified that by the fact that they would not have to read long texts, and that reading would be more difficult because the texts would be on one page and the questions would be on another. In addition, reading would require more time for concentration, as it is more difficult focus on and read from the screen, unlike grammar which depends more on previous knowledge. As for the 30% who thought that reading would be easier, the reason behind that was that they expected reading from the screen would be clearer and better organized.

However, in the post test focus group all of the students without exception replied that the grammar section was easier, because it was all multiple choice questions which required less visual focus. Secondly, it was easy to move from one question to another as well as to change the answers, unlike the reading section which required lots of visual focus in order to read passages which had an effect on sight due to long concentration on the screen.

Also most of them reported that reading from paper is much easier than reading from the screen. One student said *"Reading from paper is easier because we are used to reading from books and we can highlight what we want by using a pen and we can also write our comments in the margin"*. However, another one said that *"if I were given the freedom to choose the font size and its colour that would make it easy for me to read from the screen"*. Also another one added *"I liked the idea of doing the test through computers in the grammar section, but I did not like the reading section because of the design and layout of the test"*. Generally, students agreed that reading from paper is easy because they are used to it.

8.9.4. Language Effect

When the students were asked whether the computerised test would be easy if taken in their mother tongue, they all agreed that it would definitely be so. One student commented that "*The mother tongue would make the computerised test 90% easy*". Another one added "*It depends entirely on the student's level of English and his computer skills*".

In general most of the students said that it would be easier to take computerised assessment in their mother tongue. In addition, they believe that using computers in English would increase the difficulty, especially for foundation year students whose level in English is still relatively low.

It seems that the language has affected students' acceptance of CAA. As one of them said "*It would be good if we got used to using the computer assessment in our mother tongue and then gradually started to use it in English*". Another one also said "*having a test in our mother tongue would be easy; because of our understanding of lots of terms in it rather than in another language*".

So the "*language effect*" added another burden and caused more anxiety to students, especially as all foundation year students still do not have a fairly good level of English language. Therefore, students' anxiety and worry did not just stem from the novelty of the mode, as this interacted with their low level of English which made it even more difficult for them to deal with the computerised test; as one of students said "*if the computerised test was in my mother tongue my confidence would increase and my worries would be less*".

8.9.5. System Reliability

It became clear from the focus group discussions that students do not trust the internet connection or the computers in their institutions. Some of them said that even if they accepted CAA and wanted their exams to be by computer, the technical problems in the computers lab will make them avoid it. When I asked them to give more details about these problems they mentioned the following points:

- Many computers have viruses one student said *"I do not trust college computers since I have lost my work many times before because of viruses"*
- Lack of qualified technical staff who can help with computer problems
- Slow internet connection and sometimes the electricity failure.
- Computer technical problems.
- Unavailability of enough computers.

When asked whether they would accept CAA if all those problems were overcome, they replied by saying "yes". They also said that their trust in CAA would increase and that they would accept it without fear or hesitation because other factors can no longer affect their performance.

8.9.6. CAA Strengths and Weaknesses

When students were asked about what things they liked and the ones they disliked about the computerised test, most students replied by saying they mainly liked both accuracy and instant marking. As one student asserted *"I liked the computerised test because I got my result instantly unlike the paper pencil test where I have to wait for weeks to get my result"*. One of them also said *"I liked the idea of doing the test through computers, but I did not like the design and layout of the test in term of colours and font size."*

However, the students mentioned lots of things they liked about the computerised test such as the objective type of the questions in CAA and the fact that the computerised test would be more organised than the paper-pencil one. One student declared "*it was easy to change my answers without scratching on the paper or using the eraser, which would otherwise affect the answer paper*".

Also students mentioned other things they liked about the computerised test such as the making it easy to go through the questions and the pages. Also one of them said "*trying out a new thing that was thrilling and interesting*".

On the one hand, some students mentioned that they liked the computerised test because it prevented other students from exceeding the time set for the test, and on the other hand some students said they disliked the fact that computerised test prevented them from doing that. One student said "*I don't like time limits in the computerised test because in the paper test I could steal some more minutes if I couldn't finish on time*".

Most things disliked by the students in the computerised test were the effect of the screen on sight, especially if the test is long, as well as the substantial amount of time and concentration required for reading from the screen. One of them stated "*I could not concentrate due to the difficulty to read from the screen and its effect on sight*".

Another thing they disliked about the computerised test was being only in one colour. As one of the students said: "*It would have been very interesting if the test was in different colours.*" Another one added "*It would have been better if I could control the colours*". And a third one stressed "*It was boring that the whole* *test was in one colour"*. When I told them that the paper tests are also in one colour, they justified that by having higher expectation for CAA.

At the end of the discussions, lots of the students confirmed that there was nothing to be afraid of, also it was a fun and exciting experience. Overall, when the post-focus groups were asked to describe in one word their feelings about the experience of taking a computerised test, males used such words as: Perfect; The Best; Wonderful and Interesting, while females used the words: Good, Average, Fine..

8.10. Discussion

The most important finding in this chapter is that although there were no significant differences in students' attitudes in relation to gender, college or region, deeper and more thoughtful analysis revealed that females were more nervous than males and found some difficulty in reading from the computer compared to male participants. This difference between males and females appeared clearly through the focus group discussions. Females declared that they felt uncomfortable in completing the test by computer. One female said "*even though I did well in the computerised test, I still prefer the paper-and-pencil one*". In contrast, males' worries and anxiety decreased after sitting the computer test as they showed positive attitudes and preference for CAA. The findings might explain why males did better in the computerised test than in the paper-pencil one while females did better in the paper-pencil test than in the computerised one. This result is consistent with the findings of many studies which concluded that females feel more anxious, nervous and uncomfortable taking computerised tasks than males do (Broos, 2005; Brosnan, 1999; Graff, 2003; Shashaani & Khalili, 2001).

For example, Shashaani & Khalili (2001, p.368) assert that "females felt more helpless around computers, stating that computers made them nervous and uncomfortable". Also this result was supported by Coniam (2006) which concluded that, in Hong Kong, there was a difference between boys and girls in terms of mode preference as boys preferred computer assessment and girls were rather in favour of paper assessment. Also this finding of the present study may be explained by the fact that in Oman, males spent more time with computers inside and outside the home than females do. So it may be that as a result of computer exposure, males feel more comfortable around computers compared to females.

Moreover, the results suggest that familiarity with the computerised test is the most fundamental factor to the students. It can be concluded from the pre- and post-test focus group discussions how important the computer familiarity factor is in shaping students' attitudes towards computerised assessment. This has been made evident, particularly when students' attitudes towards CAA became more positive after taking the test. Besides, students in the focus groups declared that if they had an adequate CAA training it would have been easier for them to take computerised assessment without any fears. This result was also confirmed by the questionnaire respondents as most of the students responded by saying that they would have found helpful if they had taken a training course on computerised test. This result is supported by some studies where the majority of participants, after taking the computerised test, would prefer to be taking future test by computers (Higgins, Russel & Hoffmann, 2005; Poggio et al., 2005; Wang, Young & Brooks, 2004).

Another factor which students thought could affect their performance was the "*language effect*". While many studies talk about "mode affect" (Brown & Augustine, 2000; Choi & Tinkler, 2002; Clariana & Wallace, 2002; Dimock & Cormier, 1991; Pommerich, 2004; Wallace & Clariana, 2005) very few concentrated on "*language effect*" (Wolfe & Manalo, 2004). In fact, it has been shown that the students whose level of English is rather low would be disadvantaged by computerised assessment as argued by Wolfe & Manalo (2004) "it is likely that such an effect would be more pronounced for [the] examinees for whom English is a second language because these examinees would perform a double translation – native language to English and then English to keyboard strokes" (p.54). This issue could lead to equity issues concern. Therefore, it could be suggested that further research should target a different cohort of students whose level of English is fairly good in order to overcome the language effect factor, which might yield different results.

Also the findings of this study suggests that students tend to accept technology if the institution provided a well-designed system which students can depend on without having to worry about technical problems, which may affect their acceptance of and preference to interact with technology. However, it should be noted that most of the technical problems students have mentioned such as computer viruses, lack of qualified technical staff and the unavailability of enough computers are easy to solve. As a researcher I have faced technical problems as well. For example: computer system failure (freeze or stuck), sluggishness of network and internet connection, speed of computer because it is old and over used by students, which confirmed the students' points. Therefore, it was made sure that a bigger number of computers than needed should be available in case any problem occurs. Also, it was made sure that the researcher could get enough technical support before and during the administration of the computerised test.

The present study also shows that the students' attitudes towards computerised tests were not related to the test performance, computer experience or computer self efficacy. This result is, however, inconsistent with Comber et al., (1997) who stress that the more experienced with computers an individual is, the more positive attitudes they are likely to develop about computers, which increases their level of computer self-efficacy. The findings of this study might be explained by the fact that students' unfamiliarity with the novel assessment mode has caused them to be less confident dealing with the computerised test even though they had adequate computer experience.

Generally speaking, this study results are consistent with the findings of many other studies which have suggested that students have positive attitudes towards CAA (Sheader, Gouldsborough & Grady, 2006). Moreover, this study suggests that the acceptance and preferences of this kind of assessment method would increase if the students become more familiar with computerised assessment through CAA training courses. Also, it is important to build a technology trust through providing a more reliable system to avoid the technical problems which affect students' attitudes towards CAA.

Chapter Nine: Staff Attitudes towards Computer Assisted Assessment (CAA)

9.1. Introduction

Although many studies have investigated students' attitudes towards CAA (Broos, 2005; McKenna, 2001; Shashaani & Khalili, 2001), there are few studies that have focused on the academic staff attitudes and perceptions towards implementing CAA at university (Hodson, Saunders & Stubbs, 2002; Sheader, Gouldsborough & Grady, 2006). Because the Applied Science Colleges in Oman are planning to implement CAA, academic staff opinions and perceptions are one of the most important factors we have to explore. However, the Applied Science Colleges in Oman have staff members from over the world like all Omani higher education institutions. So this might add another layer of complexity to the situation for institutions if they wanted to implement CAA, for these staff members have different background cultures. Therefore, the purpose of this chapter is to explore staff attitudes towards CAA and to see if they vary according to gender and nationality Also to investigate staff perceptions of the difficulties and limitations which might face CAA, and to explore the advantages and disadvantages of CAA from the staff points of view. In addition, this chapter will compare staff and students' opinions and views about using the computer as an assessment tool to see if their concerns are similar or different.

Both quantitative and qualitative research methods were applied by using a questionnaire and semi-structured interviews to get detailed information about staff opinions and perceptions towards CAA. Similar statistical procedures would be used throughout this part of this study. Finally, the discussion will summarize the finding of this chapter in the light of the literature.

9.2. Staff Characteristics

The academic staff members in these colleges are from different countries and have different mother tongues, Arabic or English. Table 33 below shows the sample sizes in relation to gender and nationality/language. The percentage of the sample is representative to some extent of the population of all the Applied Science Colleges (See Table 8, p.150 in Chapter Five).

Variables		Count	Percentage
Gender	Male	86	70.5%
	Female	36	29.5%
Nationality/ language	Omani	15	12.3%
	Arabic non Omani	32	26.2%
	English non Omani	75	61.5%
Total staff sample		122	100%

Table 33. Sample Size in Relation to Gender and Nationality/language

Of the 122 academic staff who participated in the study 86 (70.5%) were males and 36 (29.5%) were females. 15 (12.3%) were Omani and 32 (26.2%) were Arabic-Speaking non-Omanis while 75 (61.5%) were English-Speaking non-Omani.

9.3. Internal Structure and Reliability of the Staff Attitude Questionnaire

The 18 items of the questionnaire which was developed for this study were subjected to principal component analysis. The Kaiser-Meyer-Oklin value was

0.85 exceeding the recommended value 0.60 (Kaiser, 1974) and the Barlett's test of sphericity reached statistical significance supporting the factorability of the correlation matrix.

Principal components analysis revealed the presence of five components with eigenvalues exceeding 1 explaining 35.04 percent, 10.70 percent, 8.13 percent, 6 79 percent, and 6.11 of the variance respectively. The total variance explained was 66.78% percent (Appendix E).

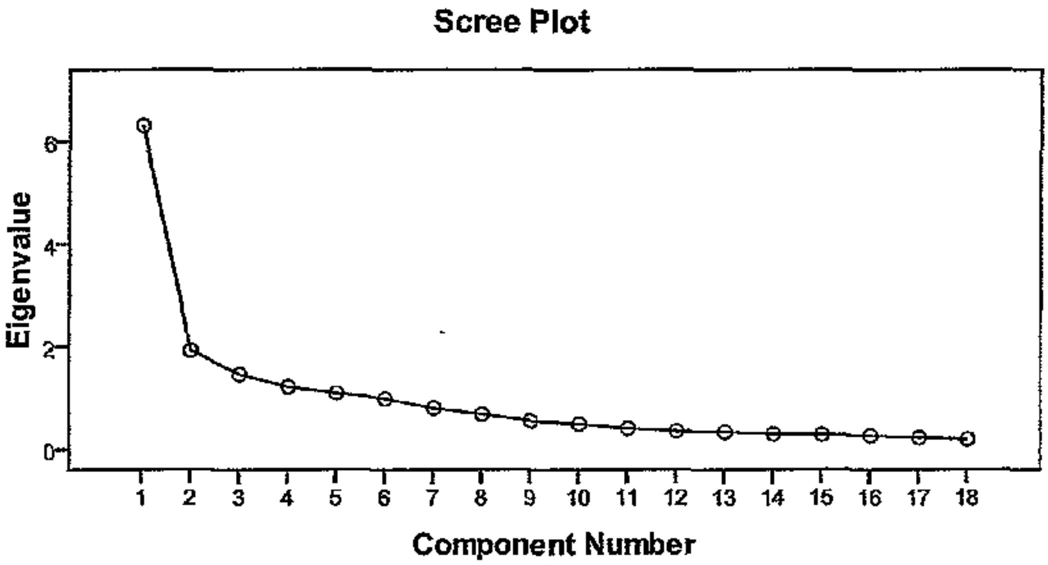


Figure 40. Scree Plot for the Staff Attitudes Questionnaire

However, the inspection of the scree plot (Figure 40) reveals a clear break after the first component. It was decided to retain one component for further investigation. The one component solution explained a total of 35.04 percent of the variance (See Appendix E). The item correlation was good, except item number 9 which was 0.29. It should be noted that the same item had a low correlation in the pilot study and this large sample confirmed it So, this item was removed from the analysis.

The internal consistency of the questionnaire was good (Cronbach's alpha = 0.88). Moreover, the item-item correlation was 0.29 to 0.78 and the Squared multiple correlations (SMC) ranged from 0.20 to 0.71. Together, these findings provide good evidence for the internal consistency of the scale.

9.4. Descriptive Statistics of Staff Attitudes towards CAA.

Table 34 below displays the mean and standard deviation (SD) of attitude scores on the 17 remaining items according to gender and nationality/ language. Five items of the questionnaire have been reversely coded to let higher scores reflect more positive attitudes towards CAA. The individual marks were summed to give score between 17 and 85

Variables		Mean	SD
Gender	Male	57.30	9.65
	Female	59.08	10.68
Nationality/language	Omani	57.47	6.73
	Arabic non Omani	60.91	8.72
	English non Omani	56.59	10.76
Mean and SD for the whole staff sample		57.83	9.95

Table 34. Attitudes Mean and Standard Deviation in Relation to the Gender and Nationality/Language

As shown in table 34 above, the mean scale score of the female staff was slightly higher than that of males. As for the nationality/language variable, it can be seen that Arabic-speaking non-Omanis have the highest mean, whereas there are similar means for the Omanis and English-speaking non-Omanis.

9.5. Test of Equality of Means (ANOVA and Multi Factor Analysis of Variance)

In order to test the hypothesis which is *Ho: There are no difference between academic staff attitudes towards Computer Assisted Assessment in regard to their gender, nationality and mother tongue language,* histograms and box plots, presented in the Appendix E, an independent t-test was conducted to determine if there is a significant difference between the mean scores of the gender variable and one way ANOVA for the nationality/language variable which has three groups. Then all these variables were combined together in a Multi Factor Analysis of Variance.

The t-test found no significant gender difference *[t (120) =0.901, p= 0.370)]*. In addition, the effect size was very small (Eta Squared =0.007).

For the nationality/language variable, one way ANOVA result was not significant as well *[f (2,119) =2.16; p=0.119]*, indicating comparable mean scores for the three nationality/language groups. Moreover, the effect size was small (Eta squared=0 035).

Multi-Factor Analysis of Variance was used to test the null hypothesis that all effect level's means are equal, and simultaneously for each of the demographic variables (see table 35 below). The assumption of normality and linearity were tested and satisfied. Staff attitudes, averaged over 17 items, do not differ in terms of males and females or Omani, Arabic-Speaking non- Omani and English-Speaking non-Omani when all explanatory factors are examined together, which confirmed the simple t-test and ANOVA results.

Tests of Between-Subjects Effects

Dependent Variable: Attitude

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Eta Squared
Corrected Model	656.552[a]	5	131.310	1.344	.251	.055
Intercept	200462.406	1	200462.406	2052.244	.000	.947
Gender	59.472	1	59.472	.609	.437	.005
National	140.094	2	70.047	.717	.490	.012
gender *national	51.098	2	25.549	.262	.770	.004
Error	11330.834	116	97.680			
Total	419963.000	122				
Corrected Total	11987.385	121				

a. R Squared = .055 (Adjusted R Squared = .014)

Table 35. Multi Factor Analysis of Variance to Gender and Nationality/Language Variables

9.6. Analysis of Item By Item Staff Questionnaire

The scale total score does not show whether there are differences in attitudes at the item level. Therefore, item by item analysis was undertaken to explore if there is any difference between males and females or between different nationality groups.

Using Chi Square tests showed that no items were significantly different in terms of males and females. So the results in the table 36 below presents the number of participants selecting agree (i.e., 'agree' or 'strongly agree') or disagree (i.e., 'strongly disagree' or 'disagree') divided by nationality. There are four of the 17 items in the scale which differ significantly in terms of Omani, Arabic- Speaking non Omani, and English-Speaking non Omani.

	Statements	National/language	Agree n(%)	Disagree n(%)	Chi square p-value
1	computer very useful	Omani	15(100)	0(0)	0.381
		Arabic	31(100)	0(0)	
		English	71(96)	3(4)	
2	like more CAA exams	Omani	4(67)	2(33)	0 130
		Arabic	20(87)	3(13)	
		English	32(64)	18(36)	
3	use computer at work	Omani	15(100)	0(0)	0.261
		Arabic	31(100)	0(0)	
		English	67(94)	4(6)	
4	computer test is easier than paper test	Omani	2(40)	3(60)	0.073
		Arabic	16(80)	4(20)	
		English	28(53)	25(47)	
5	feel confident delivering tests by computers	Omani	3(43)	4(57)	0.626
		Arabic	9(56)	7(44)	
		English	32(61.5)	20(38.5)	
6	questions will be very clear in the screen	Omani	5(56)	4(44)	0.244
		Arabic	20(83)	4(17)	
		English	26(70)	11(30)	
7	instructions will be easy in computer	Omani	5(71)	2(29)	0.814
		Arabic	15(14)	7(32)	
		English	19(61)	12(39)	
8	use computer in the house	Omani	15(100)	0(0)	0.58
		Arabic	32(100)	0(0)	
		English	63(89)	8(11)	
10	dislike assessing student by machines	Omani	8(67)	4(33)	0.054
		Arabic	21(87.5)	3(12.5)	
		English	33(60)	22(40)	
11	computerized test can only assess low cognitive ability	Omani	8(80)	2(20)	0.918
		Arabic	16(76)	5(24)	
		English	40(74)	14(26)	
12	hesitate to use computer for fear of losing all my work	Omani	13(93)	1(7)	0.158
		Arabic	25(89)	3(11)	
		English	47(76)	15(24)	
13	using CAA in the second language is difficult	Omani	12(100)	0(0)	0.084
		Arabic	19(79)	5(21)	
		English	30(70)	13(30)	
14	Computerized tests take longer to complete then paper test.	Omani	3(25)	9(75)	0.000
		Arabic	17(71)	7(29)	
		English	41(82)	9(18)	
15	feel comfortable using CAA	Omani	1(20)	4(80)	0 047
		Arabic	15(79)	4(21)	
		English	32(65)	17(35)	
16	like to attend computerized test training	Omani	15(100)	0(0)	0.038
		Arabic	26(93)	2(7)	
		English	45(78)	13(22)	
17	Prefer using computerized test than paper test	Omani	2(40)	3(60)	0.144
		Arabic	14(70)	6(30)	
		English	21(45)	26(55)	

18	CAA will save my time	Omani	8(80)	2(20)	.022
		Arabic	25(96)	1(4)	
		English	35(69)	16(31)	

Table 36. Proportions of Different Nationality Selecting Agree and Disagree Response on the Attitudes towards CAA Questionnaire

These items revealed that English-speaking non Omanis staff members are less likely to attend training sessions, and think that computerised tests take longer to complete than paper tests more than other the two groups do. Whereas Arabic-speaking non Omani staff would feel more comfortable delivering tests by computer and thought that CAA will save time more than other the two groups do.

These results may indicate that English-speaking non Omani staff had negative attitudes towards CAA more than the other groups. However, the semi-structured-interviews may clarify the picture and provide more detailed information about staff attitudes towards CAA and the reasons that lie behind these views and perceptions.

9.7. Semi -Structured Interviews with Staff

The main aim of the interviews was to explore the staff perceptions towards applying CAA in their institutions. The interviews are also meant to help me to investigate more the difficulties that might prevent them from implementing CAA, as well as the advantages that academic staff will gain from CAA and the potential disadvantages from their point of view.

I used face to face semi-structured interviews with 23 of the academic staff who teach in the three sampled colleges. The majority of them (17 out of 23) were from English language departments and the others (6) were from different departments in these colleges. I listened to the recordings of each interview several times to categorize the data according to the research questions listed in chapter five. I analyzed the data thematically and three themes emerged which are; CAA acceptance, CAA difficulties and limitation, and CAA advantages and disadvantages.

9.7.1. CAA Acceptance.

Only 18 out of 23 of the academic staff who have been interviewed owned personal computers, *and five of them replied by saying they do not have a* computer. These five members were in the English-speaking non-Omani group One of these five members said "*I do not like computers I even do not trust doing my work through computers*" and added "*I hate computers".* Most of the academic staff members reported that their computer skills are moderate to good (even those who do not have computers), except two of them who described their computer skills as excellent Five of the staff reported that they had not heard about computerized assessment before, eight of them knew about this mode of assessment but had never used it, while ten had used this type of assessment with such software as: Moodle, Blackboard and WebCT.

Before beginning with the interviewees, I gave a brief description of CAA to those who said they had never heard about it before so that I could proceed with the exploration of their views about it

Most of the interviewed staff were willing to use CAA but not at that time. As one of them said "*it would be a radical change and I would automatically accept it, but we need to be patient with both teachers and students because it would be a big change".* Another academic pointed out "*I do not oppose it, but I think* *it is so early to use it*". Another one added "*I am so excited to switch to computerized tests, but I think it should take its time and it should be applied steadily*".

Investigating this combination of willingness for the future and reluctance for the present, I tried to investigate more the reasons behind this hesitance about starting to use CAA. I know that all institutions had the Blackboard programme and each of all six applied colleges had been offered a three-day training course in that programme which I personally attended. It should be noted that the Blackboard programme trainer was from outside the institutions.

My goal was to know the reasons why they reject using CAA, which I called "*staff internal obstacles*" because the acceptance of CAA might be prevented by some obstacles posed by the staff members themselves. I categorise them as: lack of computer skills and lack of CAA knowledge.

Lack of computer skills by some academics may affect the acceptance of CAA, and consequently lead to resistance to change from the traditional method to CAA. One of the staff said "*some academic staff have never used computers in their life so how can they run CAA?*". Another staff member said "*we have some members who do not know how to switch the computer on and off*". So this might be a factor that affects the acceptance of CAA which requires good computer skills in order to be used. But it seems that this is not the main apprehensive factor, as most of the interviewed staff members had reported that their computer skills were moderate to good. But it may be computer self efficacy levels, which has affected their acceptance of CAA. However, this research could be a subject of future.

Another obstacle which might limit the acceptance of CAA is lack of knowledge about CAA and what it can offer. Few of the interviewed staff (4 out 23) thought it would be difficult to assess higher-order thinking skills by CAA and 17% of the academic staff agreed with that view during the questionnaire as they thought that computerized testing can only assess lower cognitive ability (see Appendix E). One member of staff stated *"I will not use CAA because it is not suitable for my subject and the kind of questions that I prepare for my students".* However, others believed that CAA can assess higher-order thinking skills but that requires lots of training for the academic staff on how to create good quality questions. Also, they believed that it needs time to create this kind of questions that are suitable for CAA. As one of the staff said "*CAA is like any assessment method and can measure any skills but it needs time and expertise in creating a good type of questions".* It seems that some of the staff knows about the potential of CAA and what it can offer, which might lead to gaining confidence of this new assessment mode, which could lead to greater acceptance.

In addition, some staff thought that CAA is not suitable for all streams and subjects. One of the staff member said "*CAA is not suitable for every subject so I will not use it for my students".* Also, some of the staff members believed that only objective questions can be used in CAA Those were the ones who declared that they had never used CAA. In contrast, the staff who had CAA experience declared that it can be used for many types of questions, objective and subjective.

Some staff mentioned that a number of current modules and curricula are still using traditional methods in teaching. This means that using CAA will lead to "constructive alignment" problems. Hence, assessment should not be changed, but has to be kept aligned with both the intended learning outcomes of a course or a module and the teaching practices carried out by academics, for it is otherwise unfair to teach students in one system and assess them in another (Brown, Bull & Pendlebury, 1997).

The staff questionnaire revealed that 50.8% of the staff thought that paper tests would take student longer to complete than computerised tests. To clarify this point through interviews it seems that most of the academic staff thought that the paper-pencil test would take more time than the computerized test unless the computerized test questions do not require high computer skills and the questions types are objective ones. Otherwise, the computerized test will take a longer time especially if the students have insufficient computer skills or are required to write in another language apart from their mother tongue. Others commented that it depended on various factors such as what the test intends to assess, the question types, and the students' computer skills. That is, while some students are slow in writing and fast in typing, there are other ones who are totally different. This was also confirmed in the students' focus group discussions where students had already reported that the reasons behind worrying about accepting CAA are mainly the nature of the CAA questions, the language effect, the complexity in using computer, and time pressure especially for those who are slow in using the keyboard or typing on the computer.

Most of the interviewees, except one English speaking non Omani, were willing and ready to attend any workshop or training session on this type of assessment. The one who refused said *"I do not like to attend any training session if is it optional but if the institution shifts to CAA formally then this will* *force me to attend."* The staff questionnaire result confirmed this finding as 71% of the staff responded that they are willing to attend any CAA training sessions However, they preferred that any change should be applied gradually. Staff members suggested that they should be trained first on how to operate and apply this type of assessment and stressed that the students should be prepared psychologically until they are familiarized with this new assessment mode. Above all, resources need to be catered for, including a sufficient number of computers, fast internet connection and technicians to help overcome any technical problem. This view was supported by the students, who also expressed the need for a reliable system.

Nevertheless, all of the staff members who were interviewed agreed that they preferred to start using CAA for formative assessment until both the staff and the students are fully acquainted with such a method of assessment. Then, they could move to summative assessment, possibly centralized from the ministry to all the colleges in Oman. This was also supported by the questionnaire results which showed that just 29% of the staff would not prefer to substitute computerized testing with paper-based testing, while 71% have moderate attitudes or prefer to use computerized tests. Also, just 25% would not feel confident delivering the tests by computers while 75% would feel confident or have moderate attitudes. In addition, 21% would not feel comfortable using CAA with their students, but 79% would feel comfortable or have a moderate attitude. Also, 19% do not like exams to be administered using Computer Assisted Assessment, while 81% would like that or have moderate attitudes.

Finally, when the staff members were asked about how they would feel if their institutions switched from paper-pencil tests to computerized tests, they were all excited but they said they believed it should take place gradually.

We can conclude that, although the responses of staff members in the questionnaire (58%) showed positive attitudes towards implementing CAA, most interviewees felt that the institutions were not ready enough to implement CAA as there were many difficulties that would hinder the introduction of CAA in these colleges. The staff also seemed to be willing to implement once these obstacles are overcome.

9.7.2. CAA Difficulties and Limitation

The staff questionnaire responses revealed that 55% of the academic staff thought that implementing CAA will face lots of barriers (see Appendix E) it should be noted that item was removed from analysis because of low loaded (coefficient) with other items. So I aimed to investigate the difficulties that might face implementing CAA in Oman from academic staff point of view. I also investigated the limitations that might prevent staff members from applying CAA in Omani colleges. Those difficulties have been separated into different categories such as financial; technical, academic staff difficulties (mentioned in the above section) and students' difficulties. The following section will list these difficulties as well as the suggestions provided by the academic staff on how to overcome these difficulties.

9.7.2.1. Financial Difficulties

All colleges have learning resource centres containing laboratories with many computers. However, as student numbers have increased rapidly, the number of computers has become inadequate. So one of the difficulties mentioned by staff members is the inadequacy of computer labs compared to the number of students. One of staff said *"the number of labs is not enough compared to the student classes and sometimes we couldn't find labs to use as all of them are fully booked ".* This aligns with student views (Shapter Eight).

9.7.2.2. Technical Difficulties

Staff members mentioned a number of technical problems that affect the system's reliability. The same point was mentioned by the students who focused on the importance of the system reliability to build trust in computer related tasks, especially assessment.

The difficulties mentioned by staff members included poor internet connection, frequent shutoff of the electricity, especially in Sur College, insufficient computers, poor computer quality, unreliability because of viruses which lead to losing some data and some programs As one of the staff said *"all these computers need to be changed and replaced by good ones"* Students in focus groups agreed with the above points and confirmed that if they had reliable systems that would increase their acceptance of CAA without hesitation.

Staff members said that there is also a lack of technicians to assist the academics with computer problems and to resolve any problems that might occur. There was also a lack of good technicians to train staff members on how to use the CAA. One member of staff said *"I cannot rely on the college staff technician as they are not very qualified to assist us in computer-related problems".*

Speaking about technical problems, the staff also mentioned the problem of cheating and test security. One of staff said *"how can I make sure that my test will not be hacked by a professional student on computers".* The cheating issue has been discussed in the literature and some suggestions were made on how to minimise the possibility of cheating in CAA system. Zakrzewski and Steven (2003) suggest different techniques to avoid the possibility of students copying answers from each other during the computerised assessment. For example, some CAA programmes have facilities to display questions in a random order or to display the choices in a different sequence; hence, this will reduce the possibility of looking at the nearby monitors. *"Another possibility is to invest in privacy screen placed on workstation monitors which only enable face-to-face viewing and cut out any attempts to view a workstation from one adjacent to it"* (Zakrzewski and Steven,2003, p619). Also with CAA, the academic staff need to be vigilant before, during and after the examinations and students must present their identification documents so that *"username and password have to be checked for authenticity"* (Whittington, 1999 cited in Zakrzewski and Steven, 2003, p.619).

In terms of security, the examinations and results should be secured and the security managers have to verify that access is available for only the authorised academic members. (Zakrzewski and Steven, 2003).

9.7.2.3. Student Difficulties, as Perceived By Staff

Staff members mentioned many difficulties that might face institutions in relation to their students' characteristics such as poor computer skills and low experience, especially foundation year students. As a member of staff said *"most of the foundation year students have limited knowledge of computer use* *and sometimes no knowledge".* Some staff members mentioned that foundation year students would not accept easily the computer culture and would refuse to be assessed by the computer. One member of staff asserted that "*some students will use their lack of computer experience as a reason for their failure*"

However, many of them said that students liked computers but they needed time to become well experienced in them. One academic declared "*Omani students enjoy dealing with computers like any other youth and they will accept CAA easily after getting suitable training in it*". This was supported by students views in the focus group discussion and questionnaire as most of them said they were willing to accept CAA if they undergo suitable training and if they are given ample time to familiarize themselves with it.

In addition, students in the focus groups and the interviewed staff members agreed on the issue of dealing with the computer in a different language and reported that it would represent a hurdle, especially for foundation year students. This result was supported by the staff questionnaire responses which revealed that 72% of the staff thought that using CAA in a different language would be difficult for students (See Appendix E).

9.7.2.4. Staff Suggestions for Overcoming Difficulties

Although the academic staff mentioned many difficulties, they also presented suggestions that might help to overcome these difficulties and offer a strong fundamental base to introduce CAA in Omani institutions.

First of all, staff members suggested that introducing CAA in Omani higher education institutions should be gradual through many steps as with the introduction of computer into the teaching system and curriculum. For example, the curriculum should be designed in a way which allows Computer Assisted Learning (CAL) and Computer Assisted Assessment (CAA) to be a part of it. As a result, that will change the traditional way of lectures from solely depending on text books to depending on CAL as well. Time is also needed to introduce the new assessment system to staff and students, and accommodate CAA in Omani higher education institutions which will then increase their confidence to use it and familiarise themselves with it.

Staff members confirmed that colleges have to be prepared financially, technically and psychologically and train the academic staff on how to prepare and design questions suitable for CAA, and also students on how to use CAA till they become familiar with it. They suggested starting first with formative assessment then continuous assessment and at the end move to summative assessment. This, as staff declared, will build trust in the CAA system and will give time for students to familiarize themselves with the new mode of assessment. Zakrzewski, Steven & Ricketts (2009) argued that introducing a new method of assessment at the end of the course will increase student anxiety. Therefore they suggested introducing sampled questions on the network to enable students to familiarize themselves this new kind of assessment and the different associated questions types. Also, they argued that the evaluation of formative e-assessment will "*provide information that allows revisions and improvement to be made to risk-elimination or risk-reduction procedures*" (p.444).

Moreover, the interviewed academic staff confirmed that training sessions have to be continuously run throughout the academic year and that all colleges need technicians who are able to solve the various technical problems that staff and students might face.

Some of staff suggested starting firstly with some streams and subjects and then gradually moving on to the others. It is also necessary to reflect upon the feedback on its practicality and its usefulness before the actual application takes place by conducting studies and setting pilot projects in the colleges. This would increase the awareness of its problems and limitations as well as its advantages and disadvantages before moving to the actual implementation This is like the first stage in the 'Catherine wheel' model for a web-based CBA system which was originally developed by Zakrewski & Steven (2000) and, which has five segments· Planning, Risk analysis and management, Assessment design, evolutionary development of system and Evaluation, which implies that CAA should first be piloted in a single module before being widely implemented. According to Zakrewski & Steven (2000), "*planning and risk analysis performed based on the pilot aims and objective. An objective test is then designed, written and the pilot implemented The pilot is evaluated and a decision is made at this stage whether to expand to a departmental system or not. The evaluation will therefore involve a pilot review and a feasibility study for further growth*" (p.203).

9.7.3. CAA Advantages and Disadvantages

This study aimed to know the advantages and disadvantages of CAA from staff viewpoints. When comparing any two modes of assessment it must be the advantages that overcome the disadvantages because that will increase the utility of assessment through greater acceptability. Through the interviews, academic staff members mentioned many advantages that they might gain from implementing CAA, as well as the disadvantages.

There are many CAA advantages stated in the literature (Bull & McKenna, 2004; Chalmers & McAusland, 2002; Seale, 2002) and academic staff members in this study have mentioned the same advantages as well. Most staff members verified that the one important advantages of CAA was that it will save time especially in marking, which will lead to saving staff effort and give more time for attention to learning improvement (time redistributed). Also one of the CAA advantages is the accuracy in marking and instant feedback. Students in focus groups also confirmed that and said that in CAA they will not wait for weeks as in paper test to see their result and identify their mistakes.

Some staff members also mentioned that they would be more capable than before of understanding students' writing on computers as they are already facing difficulty in reading students' hand writing. One member of staff said: *"I will understand students' answers on computer screens, unlike when they are written on the answer paper especially for those students with bad handwriting".* Other staff members mentioned that it would be easy to make statistical analysis due to instant marking and identify students' weak points and work on ways to overcome them, as well as detecting students' learning progress through each term.

Many other advantages were mentioned as well, such as the low cost, reduction of photocopying and the use of paper besides the ease to amend typing mistakes if discovered before handing the test out, unlike the paper-pencil test especially if they have been printed out and distributed. Also formative CAA enables students to test themselves, reflect upon their performance, learn from their mistakes and eventually develop their learning strategies.

Most staff members mentioned that one important CAA advantage is that by saving the time devoted to marking scripts, it increases the frequency of conducting tests over the whole academic year, which will improve students' educational level and enhance their learning, unlike the paper-pencil test which is difficult to mark, especially with the substantially increasing number of students. One of staff said "*I cannot do lots of test because I do not have time for marking it*".

The CAA disadvantages which were mentioned from members of staff were fewer than the advantages. Moreover they were focused on the difficulties they will face rather than the disadvantages of CAA. However, the disadvantages that most of them mentioned are CAA "*Front Load*", as CAA demands good quality questions that suit computerised tests to be designed, which takes lots of time in preparation.

Also some staff said that they need to communicate with the students and using CAA might limit the direct interaction between them and the students (face to face interaction). As one of the staff said "*I need to interact with my students. Does CAA offer that?*"

Other CAA disadvantages mentioned by the staff members are the difficulty to mark subjective questions by computer, and students sitting for a long time in front of the computer that could turn it into computer skill test rather than just subject knowledge.

9.8. Discussion

This study found that there is no difference in the attitudes towards CAA regarding their gender or nationality/language (Table 35, p.237). This result suggests that although staff members in Oman have different cultural backgrounds, all of them have the same attitudes towards CAA, this could help Omani higher education institutions to shift to computerized assessment if they would like.

The present study concluded that most academic staff is willing to accept the implementation of CAA in their institutions. However they suggest implementing it gradually. They also suggest that the institutions should first offer training sessions on using CAA software, especially as academics have shown they are willing to attend these training courses. However, these sessions have to be continuous throughout the year. This result was confirmed by the questionnaire responses which showed that most of the staff would like to attend computerized test training sessions. In addition, staff emphasized the importance of having a qualified technician in every college, whom they can rely on if the staff members need help with CAA software. This point was also was by students who confirmed the need for a qualified technician as well.

This study suggests that there should be a stage of training the academic staff members on how to prepare tests that are suitable for CAA, and there should be training courses on preparing objective questions that measure higher order cognitive skills before implementing CAA. These courses will increase staff members' confidence in this mode of assessment and its ability to measure the same learning outcomes as the conventional test does. Zakrzewski & Steven

(2000, p.212) emphasize the importance of a staff development programme which aims *"to explore pedagogic issues surrounding the construction and use of objective tests"*. Also, they mentioned that the *"important elements in assessment design are test specifications"* which considers the learning outcome, level of learning (Bloom Taxonomy) and different question types as well as the answers.

McKenna's study (2001) is consistent with my study finding. Staff in McKenna's study had mentioned similar points to that mentioned by academic staff in Oman and provides a deeper analysis of academic staff's perceptions of CAA. According to McKenna's study, CAA is mainly desired by academics for the following reasons:

- Promoting regular learning behaviour.
- Enabling academics to assess students on a broader scale than is usually possible with the traditional pen-and-paper testing mode.
- Saving time and particularly because of the rapidly expanding number of higher education students.
- Another reason for developing negative attitudes towards CAA is related to its incapability of assessing higher order learning and problem solving. "Participants tended to be most critical of CAA when considering question design limitations, particularly in relation to the assessment of higher order learning" (McKenna, 2001, p.313).

The rapid increase in students' numbers in Oman higher education institutions, as well around the world, has added another burden to academic staff's responsibilities. This burden has led academics to seek new mode of assessment which reduce that load and offer some advantages over the traditional mode (Bull &McKenna, 2004, Nicol, 2007). Thus, it is important to verify that CAA will not increase the staff members' work load. So this study result has found that there are many advantages of CAA from the staff members' viewpoints. Even the disadvantages and the difficulties which have been mentioned by staff members can be overcome through a number of proposals which they suggested. For example, start with some streams and subjects and then gradually move on to the others; offer training sessions on how to use CAA and how to prepare a high quality objective test which can assess higher order cognitive skills, but the most important suggestion is to start implementing CAA steadily.

The staff in this study suggested starting implementing CAA with formative assessment then gradually moving to summative. These steps will help both staff and students to familiarize themselves with the new mode of assessment (Zakrzewski, Steven & Ricketts, 2009). It should also give staff members' confidence from the benefits they will gain from technology, which may in turn encourage them to put more energy into creating the appropriate kind of questions needed to measure the same learning outcomes measured by paper tests. It might also lead them to create questions with different features which might increase students' motivation and enhance their learning.

Finally the most important of finding of this study reveals that academic staff members have positive attitudes towards CAA. This finding was supported by the interviews' as well as questionnaire's results which showed that almost half of the staff have a moderately attitudes towards CAA or would prefer to use computerized testing as they feel comfortable and confident delivering the tests by computers. However, they are still hesitant regarding the implementation of this new assessment mode and would prefer it to be done gradually rather than instantly. Although they mentioned lots of advantages that outweighed the disadvantages such as reduced marking time, reduced paper printing, greater accuracy in marking and instant feedback, most of them suggest that it would be better to introduce computerised assessment step by step.

These results might encourage introducing CAA in Oman as staff have positive attitude toward implementing CAA after accommodating it with higher education institutions.

Chapter Ten: Discussion and Conclusion

The main aim of this study was to investigate what factors might affect Omani higher education students when computerised assessment will be implemented. Although many studies have been published to investigate mode comparability, this study is different in terms of various aspects and reasons. Firstly, until recently computer use had been an uncommon practice in Oman. However, it has witnessed a rapid increase since the late 1990s' which have impacts on Omanis lives in many different aspects including the educational ones. So, many aspects and variables were investigated in order to see whether students' performance is different across administration modes. That is, the current research seeks to explore whether the testing mode differentially affects the performance of students in terms of such variables as gender, college, or region of residence. Another question is whether computer experience and computer self efficacy vary across these variables and do they affect students' performance. Students' attitudes towards computerised assessment, and academic staff members' perceptions, thoughts, views and feelings towards implementing computerised assessment have been investigated as well.

Secondly, as the assessment culture is still dominated by the traditional paper-based mode, this study aims to explore the possibility of introducing Computer-Assisted Assessment in a culture where both academic staff and students are solely used to traditional assessment modes. In addition, because the Sultanate of Oman has very few large cities and many villages, the availability of computers and particularly internet has not always been secure, both now and in the past. However, Omani students often need to attend colleges which may be situated in other areas. Hence, this study aims to assess the way in which differences in computer culture and experience might impact on students' performance in CAA. This study also aims to contribute to the current studies conducted worldwide in this area besides raising questions and issues that may be of further interest to researchers in the Computer Assisted Assessment.

The current study used validated methods to obtain sufficient data to answer the study questions effectively. I employed both quantitative and qualitative methods through selection of various instruments, including test delivered in both modes (paper-based and computerised), questionnaires, focus groups and semi-structured interviews. All instruments were validated through a separate study which provided good validity evidence for the study instruments. Multiple methods (quantitative and qualitative) were used to get the full picture of the effect of mode administration and the factors that might affect students' performance. The main reason behind employing multiple methods was to enhance interpretability through data triangulation and ensure the findings were consistent. However, some inconsistencies made it difficult to interpret all the findings of the study, which suggests further research is needed in this field, especially in the Omani context.

This study is not, however, devoid of limitations. Firstly, only three colleges from the six colleges that are located in different region which are overseen by Ministry of Higher Education were included. Secondly, Oman has nine regions of administration and this study sampled the students from four of them (AlDhahira, AlBatinah, Al Dakhiliya and AlSharqiya). Thirdly, only first year students (known as the "foundation year" students) were sampled in the study as those students, male and female, have similar educational backgrounds but different experiences with computers, with some of them having very little or no initial IT skills. Fourthly, most items in this study were multiple-choice items, because these items do not require high levels in computer skills, and the time required to respond to the test was limited to increase motivation and decrease distractions. Finally, in this study the paper version was converted into a computerized format; many studies suggest that converting will change the appearance of items in the computer (Russell, 1999). However, the computerized test in this research context is like the traditional one, except for the administration mode.

The previous four chapters were meant to answer all the study questions of the thesis. The results obtained from quantitative data instruments for both students and staff members in the preceding four chapters can be summarised as following:

- There is a small but significant mode difference favouring paper mode.
- There are gender differences in test performance favouring females; however, further analysis showed that females' performance was better in the paper test while males' performance was better in the computer test.
- There are some college and regional differences in relation to the test performance.
- No gender difference in computer experience has been detected.
- There are some college and regional differences in relation to computer experience and computer self efficacy.

- There is a positive relationship between students' computer experience and students' test performance.
- Females were nervous and felt difficult to read from computer screen more than males.
- No college or regional differences in relation to the students' attitudes towards computerised test were found.
- The students' attitudes towards computerised test is not related to their computer experience or computer self efficacy.
- There are no gender or nationality differences in relation to the staff attitudes toward CAA.

As for the qualitative data, the thematic analysis revealed several key themes for both students and staff. For students, the key themes were. *gender differences, CAA familiarity, Grammar and Reading Comprehension, language effect, system reliability, and CAA strengths and weaknesses.* As for staff, the key themes were: *CAA acceptance, CAA difficulties and limitations as well as CAA advantages and disadvantages*

In answer to the fundamental question of whether mode of delivery affects test score, a significant but small, difference in mean scores was found with students tending to perform slightly higher on the paper test compared to the computerized test It is worthy of note that in general, mode groups (paper and computer groups) did not have any difference in terms of their computer experience or computer self-efficacy. Besides, students' attitudes towards computerised testing were the same in the two groups. However, from the focus groups discussion it was clear that familiarity with computerised assessment was the main concern of all the students, and this may have affected their performance. This may be attributed to the fact that computerized testing may yield lower scores on the first occasion, and this is an effect of the novelty factor, as demonstrated by many research studies which call this "*test mode effect*" (Clariana & Wallace, 2002; Bunderson, Inouye & Olsen, 1989; Dimock & Cormier, 1991). Lack of familiarity with computerized assessment is an important factor (Bennett et al, 2008) and may have a negative impact on the students' performance even though they have good computer experience. This view is stressed by Fulcher (1999, p.291) who states that "the issue of familiarity is not new in language testing. It has always been accepted that test-takers should be familiar with the item types and mode of test delivery before taking a test". So, once all Omani students become fully familiar with computer assessment programmes, then computer familiarity may become less noticeable. This point will lead us to consider the importance of question type in computerised tests, as this study used MCQs which did not require much computer experience, so the difference between the two modes (0.32) was very small. However, if the type of question was more sophisticated (audio, videos) and required high level of computer familiarity, this would lead us to inquire if that may change the picture of computers as assessment tools in Oman. However, depending on this study result, Omani students have the required potential, as they have positive attitudes toward implementing CAA after experiencing it.

The most interesting finding of the current study is that, although females generally scored higher than males on both assessment modes, they achieved lower on the computerized test than they did in the paper one while males achieved higher on computerized test than in the paper one. However, there was no computer experience or computer self-efficacy difference in regard to the gender. Also, in general, no difference in students' attitudes was detected in relation to gender. But detailed questionnaire item analysis revealed that females were more nervous about tests on the computer and felt it difficult to read questions from the screen. In addition, it was obvious from the focus groups discussion that females had less preference for and acceptance of computerised assessment. Females declared that they felt uncomfortable in completing the test by computer. In contrast, males' worries and anxiety decreased after sitting for the computer test and consequently, they showed positive attitudes towards CAA, which might explain why males did better in the computerised test than in the paper-pencil test while females did better in the paper-pencil test than they did in the computerised one. This result was supported by other studies which have reported that females tend to feel helpless, nervous and uncomfortable carrying out computers-related tasks (Broos, 2005, Coniam, 2006; Liao, 2008; Shashaani & Khalili, 2001). The current study finding might be due to the fact that boys in Oman usually have more access to computers than girls inside or outside their homes and schools, unlike the situation in the UK. In fact, even now females cannot go to internet café shops in Oman because it is considered culturally and socially odd and even unacceptable for females to enter such places, even if they are accompanied by a male relative. However, males do not have such restrictions and are used to frequenting the internet coffee shops which are wide-spread in Muscat, the capital city of Oman, as well as some big cities and even small ones The only chance for females to use the computer is when the family has one at home. The other chance for females to use computers and internet is either at school or at university or college. So, as a result of these cultural norms, males have more opportunities than females to use computers outside the home, and this may be one possible explanation for the finding of differential performance between males and females. As a result, and based on previous research findings reporting that on average females have higher computer anxiety levels than males, it is reasonable to conclude that females may be disadvantaged from the introduction of computerized assessment in Oman. However, as concluded by Wallace & Clariana (2005), females tended to benefit from computer courses and managed at the end of the course to outperform males in the final examination computerized test. So, there is much in the literature to suggest that university students should be provided with sufficient training in Computer Assisted Assessment programmes and related software within the existing computer courses they take in order to familiarise them with novel assessment modes. This would not be difficult in Omani higher education institutions because they already have programs such as Blackboard, WebCT and Moodle that include facilities for CAA. A recent study conducted by Papastergiou (2010) which aimed to design and implement a computer literacy course for first year undergraduate students, supports this view. Papastergiou asserts that "data analysis showed that the course significantly enhanced students' computer and internet self-efficacy and their positive attitudes towards computers and the internet, while significantly reducing their computer anxiety" (p. 298). Papastergiou also concludes that the course had "more positive effects, in terms of computer self-efficacy [....] and computer attitudes" (2010, p.298).

However, it cannot be assumed that introducing computer courses in the Omani higher institution will definitely eliminate the gender gap in CAA because the cause might lie deeper in psychological or socioeconomic factors that affect females' performance and attitudes towards computerised assessment. So, more research in the field of psychological and socioeconomic factors is needed which might clarify the reasons behind females' negative feelings toward computer related tasks which can affect their performance in computerised assessment.

In relation to the college and region groups, the findings of this study reveal some college and regional differences relating to test performance. In spite of these differences on test performance, there were no differences in the students' attitudes towards computerised assessment in relation to these variables. However, there was a significant difference in students' computer experience levels according to which the colleges or regions with high computer experience were associated with higher test performance scores. As there is a positive significant relationship between computer experience and the total test performance scores, this may explain some of the differences relating to college and region test performance scores

This study has found some significant college and regional differences in relation to computer experience and computer self-efficacy, however the most surprising result is that students with higher computer experience had less computer self efficacy compared to those with lower computer experience whose computer self efficacy was higher. It is difficult to explain this result but one possible explanation is that increased experience with computers does not necessarily translate into a high level of computer self-efficacy First, because self-efficacy is based on self perceptions towards a well-defined task, for according to Bandura (1977) self-efficacy is domain-specific; this leads to considering the importance of distinguishing between possessing the requisite skills for something and having the perceived ability to deploy these skills to perform actions or attain expectations. That is, self-efficacy is not dependent on the skills which an individual may have, but it rather depends on what that individual can do with these skills. The psychological construct of self-efficacy is highlighted by Wood & Bandura (1989, p.408) who relate this concept to "beliefs in one's capabilities to mobilize the motivation, cognitive resources and courses of action needed to meet situational demands". Also, there are other studies which came to the same conclusion and noted that increased computer experience does not necessarily translate into increased computer self-efficacy (Sam, Othman & Nordin, 2005). In our case even though some students may have a high level of computer experience in general, their self-perceptions regarding computerised testing in particular may not be that high. "Because self-efficacy is based on self-perceptions regarding particular behaviours, the construct is considered to be situation specific or domain sensitive. That is, an individual may exhibit high levels of self-efficacy within one domain while exhibiting low levels within another domain" (Kurbanoglu, 2003, p. 636). Also "belief and reality do not always perfectly match. On one hand, talented people may suffer from self-doubt about capabilities they possess, on the other hand, despite possessing a modest repertoire of skills people may be confident about what they can accomplish" (Pajares, 2002 cited in Kurbanoglu, 2003, p.642).

In spite of the differences between computer experience and computer self efficacy levels, this study did not find any relationship between them (computer experience and computer self efficacy) and students' attitudes towards computerised assessment, which is inconsistent with the findings of other studies which have found that attitudes towards computers are related to computer experience (Chen, 1985; Shashaani, 1997). However, the samples in some studies in the literature posed a high level of computer experience which may differ about this study sample which has other characteristics. For example, some Omani students like to deal with computers, but because there may be a number of reasons that prevent them from dealing with computer till they reach the university. For example, some families do not possess computers and live in remote places with no internet café near them Therefore, if computerized assessment is implemented in all colleges which are situated in different regions, it cannot be assumed that all students would have the same level of computer experience, confidence or attitudes, for there would be regional differences that could affect students' performance. So, CAA should be designed to deal with a wide range of students with mixed abilities in terms of computer skills so as not to overwhelm low experienced students or favour those with higher experience levels over their peers which will lead to equity and validity issue. These subgroup differences in terms of gender, college and region in test performance may be a great threat to mode fairness and be a source of construct irrelevant variance, so additional research in this field is still needed.

It is worth noting that most of the staff seemed willing to adopt and use CAA, however, it was made clear that they need time to familiarize themselves with CAA programs. Staff members also suggested that introducing CAA in Omani higher institutions should be carried out gradually through many steps starting with including computers in teaching system and curricula. Also, staff members confirmed that Omani higher education institutions have to be prepared financially and technically before the implementation of CAA, and that both staff and students need to be prepared psychologically for such implementation. Staff members should also be trained on how to prepare and design questions suitable for CAA while students should be trained on how to deal with it. They even suggested starting first with formative assessment then continuous assessment and at the end move on to the final or summative assessment. It seems that these suggestions are similar to the steps that were mentioned in the literature about those who seek to introduce CAA in their institutions such as Zakrzewski, Steven & Ricketts (2009) who suggested introducing sample questions to become available for students on the network to enable students to familiarize themselves with this new kind of assessment and its different questions types. Also they argued that the evaluation of formative e-assessment will "*provide information that allows revisions and improvement to be made to risk-elimination or risk-reduction procedures*" (p.444). In addition, Stephens, Bull & Wade (1998) "*have identified several key recommendations for those who are seeking to introduce CAA:*

- *Establish a co-ordinated CAA management policy for the CAA modules and each discipline on campus;*
- *Establish a CAA unit;*
- *Appoint discipline co-ordinators within departments;*
- *Establish CAA discipline groups;*
- *Provide funding;*
- *Organize staff development programmes;*

- *Establish validation procedures;*
- *Identify technical issues;*
- *Establish operational and administrative procedures" (cited in Hodson, Saunders and Stubbs, 2002, 147).*

It is important to note that student focus groups and staff interviews gave a clear picture of the quantitative findings and explained some of the results. For example, it was clear from students' and staff interviewees that familiarity with CAA software is a key factor of the acceptance this new assessment mode. Moreover, students and staff have yielded common points in relation to CAA implementation like the need for reliable computers that staff and students can depend upon, the lack of qualified technical staff who can support staff and students professionally, computers with technical problems and the shortage in the number of computers compared to students' numbers, and the reliability of internet, particularly as most students and staff are constantly complaining about the sluggishness of internet in their institutions. I confirmed these problems mentioned by staff and students as I faced some of them while running the computerised test either in validation study or in the main study Also my findings are supported by a recent study conducted by Al-Senaidi, Lin & Poirot (2009) Their study investigates the barriers that might face Omani higher education institutions from adopting technology in teaching and learning. Their finding showed five key factors that academic staff member believe as barriers in applying ICT to their teaching practices. These factors are "lack of equipment, lack of institutional support, disbelief of ICT benefits, lack of confidence, and lack of time" (p.575).

Students and staff agreed on many points about the advantages of CAA. These benefits include accuracy and fairness in marking, as well as saving staff time, and also enhancing student learning by increasing formative self-assessment. CAA would also enable staff members (as the staff mentioned) to improve and refine their teaching strategies depending on the feedback provided by CAA about students' levels and their potential weaknesses. The most important point that both students and staff agreed on and asserted that implementing CAA should be carried out step by step till all parties are fully familiarized with the new mode of assessment and trust in the CAA system is completely established.

In regard to "*Language effect*", students in the focus group have mentioned that CAA in English language has added another difficulty to the test as their English level is low. The staff confirmed this point during the staff questionnaire as most of the staff confirmed that using CAA in a second language may render the process more difficult for the students, especially the foundation year students, as it also clarified by the staff interviews. However, this difficulty may decrease since the English language proficiency is improving through the foundation year.

So, is it the right time to establish CAA in Oman? The answer to this question should be based on the results of deep empirical research which has to investigate the application of CAA from different perspectives. It should also be noted that the result of this current research study may well trigger new theoretical and practical ideas and encourage Omani researchers to make use of the advancement in computer technology, and possibly facilitate or encourage the adoption and implementation of a different assessment mode.

Today, using Information Technology (IT) in education is a hot issue in Oman as all educational institutions have embraced it on a large scale. Moreover, many people relate the use of IT in Computer Assisted Learning to looking for information on the internet. Al-Musawi & Abdelraheem (2004) state that

> "In 2001 Sultan Qaboos University began the implement e- learning using Web CT. At the beginning there were only 8 online running courses and 981 users By the end of autumn 2002, 40 running courses were offered to different colleges at SQU with 33.001 students enrolled. Nowadays, IT is used widely in higher education in Oman. Most students easily navigate the internet using emails and searching for knowledge resources" (p.364).

As for assessment, it should be noted that due to the rapidly increasing number of students in colleges and universities (which is a current phenomenon not only in Oman but also in many other countries), assessing each and every student puts much pressure on busy academics And as the computer labs which are available in colleges and universities are already connected to the internet, then, it seems that considering the possibility of applying CAA would be not only a feasible but also a practical idea to improve the assessment system in Omani higher education institutions In other words, it seems that it is now high time the benefits of CAA were seriously considered to improve the assessment of students for both formative and summative purposes.

It has been pointed out by staff that applying CAA in higher education would save time. So, it would enable academics to better utilize their time and accordingly devote their effort to planning and devising more carefully planned and frequently administered tests as well as improving their teaching activities, which enhances students' learning. Another point is that applying this new assessment (CAA) may enable colleges and universities to get a clearer picture and to better monitor the achievement of students, and this would facilitate remediation by facilitating the process of taking practical decisions on resources allocation and future direction (Leung, 1998).

However, despite the numerous advantages of applying CAA mentioned by staff or student in the current study or from the literature, changing any assessment system in any country is not an easy process, for assessment is closely related to the socio-cultural context in which it is carried out. Furthermore, Omani society has its own specificities which intervene in the application of the new assessment system which should be considered and taken into account when dealing with the Omani student population. This could complete the whole picture of the factors impact the implementing CAA in Omani context.

Introducing any new assessment system or mode is likely to face resistance (and sometimes even rejection) either from students, academics and administrations or from decision-takers as everyone would have his / her own views and justifications. Nevertheless, providing sufficient and strong reasons would be quite effective if a new assessment system is to be introduced and applied. The current study results give evidence in students' case as most of them, before sitting a computerised test and discovering its potential, opt to choose paper test whereas after experiencing it more than half change to computerised test if they were given the chance to choose. The students justified that by its accuracy in marking and the other benefits lying behind the application of this new assessment system that overcome the traditional method.

The acceptance factor is a fundamental issue in getting people to apply any new way of assessment Furthermore, cost is also important as it is one of the variables that affect the decision on using CAA because technology is quite costly, which means that any decision is taken according to the expected benefits and is also dependent on whether it will provide us with sufficient reasons and justifications to change the already known and applied traditional assessment system. Nonetheless, if a decision is made about the need to adopt CAA, it should first be approved by the stakeholders who might have different views and opinions about this issue depending on their previous education experiences or perhaps have some reservations which they might have acquired after reading the research.

However, when assessing the feasibility, practicality and utility of CAA, we should also evaluate the long term benefits that we could get when applying this new assessment system That is, we should consider not only the price factor but also such long run benefits as establishing an accurate, reliable and a highly valid assessment system, which helps to create an efficient educational system.

Chalmers & McAusland (2002) provide several practical tips on how to proceed with the implementation of a computer based assessment system in any educational institute. These tips are presented below which Adopted from Chalmers & McAusland (2002, p.28)

- Design the aims and objectives of the project for specific pedagogic reasons, appropriate to the conditions in your department and university

Ask how it will enhance students' learning, rather than how much time or resources it will save. At the same time, CAA, though it may cut the marking load overall, rarely saves, and often adds to, the developer or co-coordinator's time.

- Make sure that your aims are achievable given the resource constraints and support available-It is better to do it well on a small scale at first, than badly on a large scale all at once.
- Conduct meaningful evaluations measuring how well the objectives have been achieved in the appropriate timescale. Verbal feedback through focus groups, for example, can be tremendously useful – tutors should be flexible enough to respond to feedback both during and after the implementation.
- Do not forget to explain to your students why CAA is being implemented and at the same time involve them in it. This way they will share in the project's objectives.

Concerning the questions set at the beginning, this research study result has provided some practical answers through investigating the impact of applying such a new assessment mode on those who would be directly affected by the application of such a process. We can conclude that students and staff in Omani higher education institutions are willing to accept implementing CAA and interact with this new assessment, however, they need time to familiarise themselves with this new method . Therefore, this study suggests many steps in order to increase the confidence and familiarising students with e-assessment. For example, as all colleges require that all graduates should attend a computer literacy course, it may be worthwhile to examine the curricula of such courses and modify the content to include the commonly required prerequisite skills of Computer Assisted Assessment (CAA) Also, technical support should be made available to help with both hardware and software problems. Besides, follow up interviews and surveys should be conducted to determine overall satisfaction on CAA As most higher education institutions already use commercial products for their courses, a more detailed analysis addressing the use of these popular systems should be made to determine the specific computer skills required and decide on their suitability for the current, relatively shallow students' computer skills levels

Finally this study is significant because it has contributed to the existing body of literature in the field of Computer Assisted Assessment from the specific Omani cultural context. The inconsistencies in findings between this study and other studies may result from the contextual situation. Moreover, this study emphasise on the importance of providing CAA training for students within an existing computer courses in the Omani higher education institutions Also provide proper training for the staff on how to prepare and design questions which are suitable for CAA, until both students and staff became familiar with CAA programs.

So, CAA should be implemented gradually starting with formative assessment to continuous assessment ending with summative assessment Also it is important to support the higher education institutions financially and technically before implementing the CAA In addition, both students and staff need to be prepared psychologically to assimilate and accommodate CAA. Furthermore, CAA programs should also deal with wide mixed abilities in term of computer experience and other variables to ensure the equity between subgroups.

However, further research is needed to explore other variables that might affect using computers as assessment tool in Omani higher education institutions. To conclude, as using technology in teaching and learning now days is a hot issue in Omani higher education institutions, it seems that it is necessary to begin to align the process with assessment.

References

A'Sadoun, H (1998) 'Teaching Computer and Computerising Education.' *Education Magazine*, 26.

Akdemir, O & Oguz, A. (2008) 'Computer-based testing: An alternative for the assessment of Turkish undergraduate students'. *Computers & Education*, 51 pp 1198-1204.

Akinyemi, A. & Al-Musawi, A (2002) 'Re-orienting instructional development practice in higher education: a case study of Sultan Qaboos University'. *Muscat, Sultan Qaboos University*,

Al-Alawi, H. (1997) *Students' perceptions in identifying and dealing with their concern about quality in the Colleges of Education-Oman*. Bath University.

Al-Alawi, H. (2004) *Issues and challenges in student assessment: the case of education colleges in Oman*. University of Technology, Sydney.

Al-Hajri, A. (2005) *Achievement tests Construction*. Muscat, Oman

Al-Hanaie, W (2005) *The effectiveness of teaching with the aid of computers in the achievement of grade 10 students in Geography*. College of Education,Sultan Qaboos University, Oman.

Al-Kother, H. (1999) 'Consruct an Arabic copy to mechanic ability test by computer' *Educational Journal*, 13 (52). pp 15-45.

Al-Lamki, S. (2002) 'Higher Education in the Sultanate of Oman: the challenge of access, equity and privatization'. *Journal of Higher Education Policy and Management*, 24 (1). pp 75-86.

Al-Lamki, S. (2006) 'The development of private higher education in the sultanate of Oman: perception and analysis.' *International Journal of private Education*, (1).

Al-Musawi, A & Abdelraheem, A. (2004) 'E-learning at Sultan Qaboos University: status and future'. *British Journal of Educational Technology*, 35 (3). pp 363-367

Al-Rabiey, S. (2002) *Implementation of instructional technology in the colleges of education in Oman*. University of Manchester

Al-Rawas, I. (2001) ' The role of the university in the preservation and further development of Omani culture', *Paper presented at the University of the 21st century international conference*. Muscat

Al-Saleem, N. E (2006) *An investigation into beliefs about information and communication technology held by faculty members in an Omani university and their professional practice with information and communication technology* University of Exeter.

Al-Senaidi, S., Lin, L. & Piroter, J. (2009) 'barriers to adopting technology for teaching and learning in Oman'. *Computer & Education*, 53 pp 575-590

American Educational Research Association, American Psychological Association & National Council on Measurement in Education (1999) *Standards for educational and psychological testing.* National Council on Measurement in Education.

American Psychological Association (1986) *Guidelines for Computer-Based Tests and Interpretations.* Washington, DC: Amer Psychological Assn

Angelo, T. A. (1995) 'Reassessing (and defining) assessment'. *AAHE BULLETIN*, 48 pp 7-9.

Askar, P. & Davenport, D. (2009) 'An investigation of factors related to self-efficacy for Java Programming among Engineering Students'. *The Turkish Online Journal of Educational Technology*, 8 (1), (Accessed: 2010 from www.tojet.net/articles/813.pdf).

Austin, J. & Mahlman, R. (2000) 'Using the Internet for career–technical assessment: A pilot project in Ohio'. *Journal of Career and Technical Education*, 16 (2), pp. 27–39 (Accessed: 2007, from http://scholar.lib.vt.edu/ejournals/JCTE/v16n2/austin.html).

Baggott, G. & Rayne, R. (2001) 'Learning support for mature, part-time, evening students: providing feedback via frequent, computer-based assessments', Danson, M. and Eabry, C. eds.). *5th International Computer Assisted Assessment Conference.* Loughborough University Learning and Teaching Development, Loughborough University.

Bandura, A. (1977) 'Self-efficacy: Toward a unifying theory of behavioral change'. *Psychological Review*, 84 pp 191-215

Bandura, A. (1986) *Social foundations of thought and action: A social cognitive theory.* Prentice-Hall Englewood Cliffs, NJ.

Bandura, A. & Wood, R. (1989) 'Effect of perceived controllability and performance standards on self-regulation of complex decision making'. *Journal of Personality and Social Psychology*, 56 (5), pp. 805-814 (Accessed: 2007 from www.des.emory.edu/mfp/Bandura1989JPSP2.pdf).

Barnett, R. (1995) *Improving Higher Education: Total Quality Care.* Buckingham: Society for Research in Higher Education and Open University Press.

Bennett, R. (1998) 'Reinventing assessment: Speculations on the future of large-scale educational assessment'. *Policy Information Center, Educational Testing Service, Princeton, NJ*,

Bennett, R. (2002) 'Using electronic assessment to measure student performance'. *The State Education Standard, Washington, DC: National State Boards of Education*, 1 pp 2005.

Bergstrom, B. (1992) 'Ability Measure Equivalence of Computer Adaptive and Pencil and Paper Tests: A Research Synthesis', *The Annual Meeting of the American Educational Research Association* San Francisco, A.

Betz, N. & Hackett, G. (1981) 'The relationship of career-related self-efficacy expectations to perceived career options in college women and men'. *Journal of Counselling Psychology*, 28 (5). pp 399-410.

Beverly, J., Beverly, E., Clarke, M & White, S. (2001) 'Changing perceptions in language learning and testing', *5th International Computer Assisted Assessment Conference.* Loughborough University.

Biggs, J. (1999) *Teaching for quality learning at university,.* Open Univ. Press

Biggs, J. (2003) 'Aligning teaching and assessing to course objectives'. *Teaching and Learning in Higher Education New Trends and Innovations,* pp 13-17.

Black, P. (2000) 'Research and the development of educational assessment'. *Oxford review of education,* 26 (3). pp 407-419

Black, P. & Wiliam, D. (1998) 'Assessment and classroom learning'. *Assessment in Education: principles, policy & practice, 5 (1). pp 7-74.*

Boo, J. & Vispoel, W. (1998) 'Computer versus paper-pencil assessment of educational development: Score comparability and examinee preference', *The Annual Meeting of the National Council on Measurement in Education.* Montreal, Quebec, Canada

Boud, D. (1995) *Enhancing learning through self assessment.* Routledge.

Bridgeman, B., Bejar, I. & Friedman, D. (1999) 'Fairness issues in a computer-based architectural licensure examination'. *Computers in Human Behavior,* 15 pp 419-440.

Broadfoot, P. (2000) 'Assessment and intuition'. *The intuitive practitioner,* pp 199-219

Broadfoot, P. (2001) 'Culture, learning and comparative education:' *Editorial. Comparative Education,* 3 (37). pp 261-266.

Broos, A. (2005) 'Gender and information and communication technologies (ICT) anxiety: male self-assurance and female hesitation'. *CyberPsychology & Behavior,* 8 (1) pp 21-31

Brosnan, D., Scheeres, H. & Slade, D. (1999) 'Cross-culture Training in the Workplace'. *Understanding Adult Education and Training (2nd ed.) St. Leonards, N.S.W Allen & Unwin ,*

Brosnan, M (1999) 'Computer anxiety in students: should computer-based assessment be used at all?'. *Computer-assisted assessment in higher education(Staff and Educational Development Series)*

Brown, G., Bull, J. & Pendlebury, M. (1997) *Assessing student learning in higher education.* London ; New York: Routledge.

Brown, P. & Augustine, A. (2000) 'Summary of the Findings of the 1999-2000 Screen Reading Field Test Inclusive Comprehensive Assessment System'. (Accessed: 2008 from http://dspace.udel.edu:8080/dspace/handle/19716/2329).

Brown, S. (2004) 'Assessment for learning'. *Learning and Teaching in Higher Education,* 1 pp 81-89.

Brown, S , Race, P & Bull, J. (1999) *Computer-assisted assessment in higher education.* Routledge.

Brown, S., Race, P. & Smith, p. (1996) *500 Tips on Assessment.* London: Kogan Page.

Brown, S., Rust, C. & Gibbs, G. (1994) *Strategies for diversifying assessment in higher education.* Oxford: Oxford Centre for Staff Development.

Bugbee Jr, A. (1996) 'The Equivalence of Paper-and-Pencil and Computer-Based Testing'. *Journal of research on computing in education*, 28 (3). pp 282-299.

Bull, J. (1999a) 'Computer-assisted assessment: impact on higher education institutions'. *Educational Technology & Society*, 2 (3). pp 123-126.

Bull, J. (1999b) 'Update on the National TLTP3 Project: The implementation and evaluation of computer-assisted assessment', M, D. (ed. *3rd International CAA Conference*. Loughborough University.

Bull, J., Danson, M., Conole, G., Davis, H., White, S. & Sclater, N. (2002) 'Rethinking assessment through learning technologies'. *Proceedings of Winds of Change in the Sea of Learning: Charting the Course of Digital Learning (ASCILITE 2002) UNITEC, Institute of Technology, Auckland, New Zealand.*,

Bull, J. & McKenna, C. (2004) *Blueprint for Computer-Assisted Assessment* London: Routledge Falmer

Bunderson, C., Inouye, D. & Olsen, J. (1989) 'The four generations of computerized educational measurement'. *Educational measurement*, 3 pp 367-407.

Burke, M., Normand, J. & Raju, N. (1987) 'Examinee attitudes toward computer-administered ability testing'. *Computers in Human Behavior*, 3 (2). pp 95-107.

Burton, D. & Bartlett, S. (2005) *Practitioner research for teachers.* Sage Publications Ltd.

Busch, T. (1995) 'Gender differences in self-efficacy and attitude toward computers'. *Journal of Education Computing Research*, 12 pp 147-158.

CAA Centre (2002) 'Objective Test Design', (Accessed: 2007 from www.caacentre.ac.uk›Resources/objective_test/index.shtml).

Carmines, E. & Zeller, R. (1979) *Reliability and validity assessment.* California: Sage Publications, Inc.

Cerillo, T. & Davis, J. (2004) 'Comparison of paper-based and computer based administrations of high-stakes, high-school graduation tests', *Annual Meeting of the American Education Research Association.* San Diego.

Chalmers, D. & McAusland, W. (2002) 'Computer-assisted assessment'. *The Handbook for Economics Lecturers: Assessment. Edited by J. Houston and D. Whigham. Economics LTSN, online at www.economics.ltsn.ac.uk/handbook*, (Accessed: 2007).

Charman, D & Elmes, A (1998) *Formative assessment in a basic geographical I statistics module*. eds. CHARMAN, D. and ELMES, A., Computer Based Assessment(Volume 2): Case studies in Science and Computing. Plymouth: SEED Publications, University of Plymouth.

Chen, M. (1985) 'Gender and computers: The beneficial effects of experience on attitudes', *the Conference on Computers and Children*. Ann Arbor, Michigan.

Choi, S. & Tinkler, T. (2002) 'Evaluating comparability of paper-and-pencil and computer-based assessment in a K-12 setting', *Annual Meeting of the National Council on Measurement in Education* New Orleans, A.

Chou, H. (2000) 'The Effects of Training Method and Individual Differences on Learning Performance and Computer Self-Efficacy in WWW Design Training', *The Annual Meeting of the American Educational Research Association (April 24-28)*. New Orleans, LA

Clariana, R. & Wallace, P. (2002) 'Paper-based versus computer-based assessment: Key factors associated with the test mode effect'. *British Journal of Educational Technology*, 33 (5) pp 593-602.

Clark, C. (2003) 'Towards a unified e-learning strategy: Consultation document Nottingham'. *DfES Publications*, 1 (Accessed. 2007).

Cohen, J. (1988) *Statistical power analysis for the behavioral sciences*. 2nd edn Hillsdale, N.J.: L Erlbaum Associates

Collins Dictionary (2001) *Collins Cobuild Dictionary* Glasgow, Scotland. HarperCollins

Comber, C., Colley, A., Hargreaves, D. & Dorn, L (1997) 'The effects of age, gender and computer experience upon computer attitudes'. *Educational Research*, 39 (2). pp 123-133

Compeau, D. & Higgins, C (1995) 'Computer self-efficacy: Development of a measure and initial test'. *Mis Quarterly*, 19 (2). pp 189-211.

Coniam, D. (2006) 'Evaluating computer-based and paper-based versions of an English-language listening test' *ReCALL*, 18 (02). pp 193-211.

Creswell, J. & Miller, D. (2000) 'Determining validity in qualitative research'. *Theory into practice*, 39 (3) pp 124-130.

Crisp, V. & Ward, C. (2008) 'The development of a formative scenario-based computer assisted assessment tool in psychology for teachers: The PePCAA project' *Computers & Education*, 50 pp 1509–1526

Davidson, P (2003) 'Why technology had only a minimal impact on testing in education.', *The 2nd Education Technology Conference and Exhibition*. Sultan Qaboos University, Oman.

Davis-Case, D. (1990) 'The community's toolbox: the idea, methods and tools for participatory assessment, monitoring and evaluation in community forestry'. *Community Forestry Field Manual*, 2 (Accessed. 2007 from www.Fao.org18-1-2010).

Delcourt, M. & Kinzie, M. (1993) 'Computer Technologies in Teacher Education: The Measurement of Attitudes and Self-Efficacy'. *Journal of Research and Development in Education*, 27 (1). pp 35-41.

Dimock, P. & Cormier, P. (1991) 'The Effects of Format Differences and Computer Experience on Performance and Anxiety on a Computer-Administered Test'. *Measurement and Evaluation in Counseling and Development*, 24 (3). pp 119-126.

Dochy, F., Segers, M., Gijbels, D. & Struyven, K. (2007) 'Assessment engineering: breaking down barriers between teaching and learning, and assessment'. *Rethinking Assessment in Higher Education: Learning for the longer term, Abigndon: Routledge*, pp 87-100.

Drasgow, F. & Olson-Buchanan, J. (1999) *Innovations in computerized assessment.* New Jersey: Lawrence Erlbaum.

El-Shibiny, M. (1995) 'Higher education in Oman: its development and prospects', Shaw, K. (ed. *Higher education in the Gulf: problems and prospects.* University of Exeter.

Elton, L. & Johnston, B. (2002) *Assessment in universities: a critical review of research.* LTSN Generic Centre.

Erwin, T. (1991) 'Assessing Student Learning and Development: A Guide to the Principles, Goals, and Methods of Determining College Outcomes'.

Evans, L., Tannehill, R. & Martin, S. (1995) 'Children's reading skills: A comparison of traditional and computerized assessment'. *Behavior Research Methods Instruments and Computers*, 27 pp 162-165.

Fagan, M., Neill, S. & Wooldridge, B. (2004) 'An empirical investigation into the relationship between computer self-efficacy, Anxiety, Experience, Support and usage'. *Journal of Computer information Systems Winter 2003-2004*, 44 (2). pp 95-104.

Falchikov, N. (ed.) (1995) *Improving feedback to and from students.* Assessment for learning in higher education. London: Kogan Page. 157-166 pp.

Fan ovi ová, J. & Prokop, P. (2008) 'Students' attitudes toward computer use in Slovakia'. 4 (3), pp. 255-262 (Accessed: 2010 from www.ejmste.com).

Fulcher, G. (1999) 'Computerizing an English language placement test'. *ELT journal*, 53 (4). pp 289-299.

Gallagher, A., Bridgeman, B. & Cahalan, C. (2002) 'The effect of computer-based tests on racial-ethnic and gender groups'. *Journal of Educational Measurement*, 39 (2), pp. 133-147 (Accessed: May, 2007 from www.ets.org/Media/Research/pdf/RR-00-08-Gallagher.pdf).

Gibbs, G. & Simpson, C. (2003) 'Measuring the response of students to assessment: the Assessment Experience Questionnaire', *11th International Improving Student Learning Symposium*. Hinkley, pp. 1-12.

Golafshani, N. (2003) 'Understanding reliability and validity in qualitative research'. *The Qualitative Report*, 8 (4). pp 597-607.

Goldberg, A & Pedulla, J. (2002) 'Performance differences according to test mode and computer familiarity on a practice Graduate Record Exam'. *Educational and Psychological Measurement*, 62 (6). pp 1053-1067

Graff, M. (2003) 'Cognitive style and attitudes towards using online learning and assessment methods'. *Electronic Journal of eLearning*, 1 (1), (Accessed: 2008 from www ejel.org/volume-1-issue-1/issue1-art3-graff pdf)

Guba, E. & Lincoln, Y. (1981) *Effective evaluation:[improving the usefulness of evaluation results through responsive and naturalistic approaches].* San Francisco: Jossey-Bass

Hackett, G. (1985) 'Role of mathematics self-efficacy in the choice of math-related majors of college women and men: A path analysis'. *Journal of Counseling Psychology*, 32 (1) pp 47-56.

Harris, D. & Bell, C. (1990) *Evaluating and assessing for learning.* London: Kogan Page.

Higgins, J., Russel, M. & Hoffmann, T. (2005) 'Examining the effect of computer-based passage presentation on reading test performance' *Journal of Technology,Learning,and Assessment*, 3 (4).

Hodson, P., Saunders,.D. & Stubbs, G. (2002) 'Computer-assisted assessment: Staff viewpoints on its introduction within a new university' *Innovations in Education and teaching International*, 39 (2). pp 145-152.

Holden, R & Hickman, D. (1987) 'Computerized versus standard administration of the Jenkins Activity Survey (Form T)'. *Journal of human stress*, 13 (4). pp 175-179.

Horkay, N., Bennett, R., Allen, N , Kaplan, B. & Yan, F. (2006) 'Does it matter if I take my writing test on computer? An empirical Study of mode effects in NAEP' *Journal of Technology, Learning, and Assessment*, 5 (2).

Horne, J. (2007) 'Gender differences in computerised and conventional educational test' *Journal of Computer Assisted Learning*, 23 (1). pp 47-55

Houtz, L. & Gupta, U. (2001) 'Nebraska High School Students'. *Journal of research on computing in education*, 33 (3), pp. 316-327 (Accessed: 2010 from http://206.58.233 20/jrte/33/3/abstracts/houtz html).

http://www.internetworldstats com/me/om.htm ' '

Isiksal, M. & Askar, P. (2005) 'The effect of spreadsheet and dynamic geometry software on the achievement and self-efficacy of 7th-grade students'. *Educational Research*, 47 (3). pp 333-350

Johnson, D., Ferguson, J. & Lester, M. (1999) 'Computer experiences, self-efficacy and knowledge of students enrolled in introductory university agriculture courses'. *Journal of agricultural education*, 40 pp 28-37.

Kaiser, H (1974) 'An index of factorial simplicity'. *Psychometrika*, 39 (1). pp 31-36.

Kapes , J. & vansickle, T. (1992) 'Comparing paper and computer-based versions of the Harrington-O'shea career Decision-making system'. *Measurement and Evaluation in counseling and Development*, 25 pp 5-13.

Katz, L. & Dalby, J. (1981) 'Computer and manual administration of the Eysenck Personality Inventory'. *Journal of Clinical Psychology*, 37 (3). pp 586-588.

Khorrami-Arani, O. (2001) 'Researching computer self-efficacy'. *International Education Journal*, 2 (4), pp. 17-25 (Accessed: 2008 from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.94.1197&rep).

Kies, S., Williams, B. & Freund, G. (2006) 'Gender plays no role in student ability to perform on computer-based examinations'. *BMC Medical Education*, 6 (1). pp 57.

Kingston, N. (2009) 'Comparability of Computer-and Paper-Administered Multiple-Choice Tests for K–12 Populations: A Synthesis'. *Applied Measurement in Education*, 22 (1). pp 22-37.

Krueger, R. (1988) *Focus groups: A practical guide for applied research.* Newbury Park, CA Sage.

Kurbanoglu, S. (2003) 'Self-efficacy: a concept closely linked to information literacy and lifelong learning'. *Journal of Documentation*, 59 pp. 635-646 (Accessed: 2008 from http://www.emeraldinsight.com/0022-418html).

Kurbanoglu, S. (2009) 'Self-Efficacy: An Alternative Approach to the Evaluation of Information Literacy'. (Accessed: 2010 from www.isast.org).

Lam, P. & Tham-Ng, C. (1996) 'Estimating Proficiency Change Using Computer Administered and Paper-and-Pencil Tests', *Paper presented at the Joint Conference of the Educational Research Association.* Singapore Retrieved in October 2008 from www.aare.edu.au/96pap/thamc96609.txt.

Larsen, D., Butler, A. & Roediger III, H. (2008) 'Test-enhanced learning in medical education'. *Medical Education*, 42 (10). pp 959-966.

Larson, J. (1987) 'Computer- Assisted Language Testing:Is It Profitable? '. *ADFL bulletin*, 18 (2). pp 20-24.

Lawson, D. (1999) 'Formative assessment using computer-aided assessment'. *Teaching mathematics and its applications*, 18 (4). pp 155-158.

Lee, G. & Weerakoon, P. (2001) 'The role of computer assessment in health professional education: A comparison of student performance in computer-based and paper-and-pen multiple-choice tests'. *Medical teacher*, 23 (2). pp 152-163.

Leung, C. (1998) 'Computerized adaptive testing as a means for mathematics assessment'. *EduMath*, 7 pp. 21-27 (Accessed: July 2009 from www.fed.cuhk.edu.hk/~fllee/mathfor/edumath/9812/05leungck.html).

Liao, Y. (2008) 'Does gender gap on computer anxiety exist? A meta-analysis of studies from 1984 to 2007'. *TECHNOLOGY AND TEACHER EDUCATION ANNUAL*, 19 (2). pp 1165-1173.

LTSN (2002) 'The Learning and Teaching Support Network (LTSN) Generic Centre e-learning and assessment guides', (Accessed: August 2007 from http://www.ltsn.ac.uk/genericcentre/index).

Mackenzie, D., O'Hare, D., Paul, C , Boyle, A , Edwards, D., Williams, D. & Wilkins, H. (2004) 'Assessment for Learning: The TRIADS . Assessment for Learning Outcomes Project and the development of a pedagogically friendly computer-based assessment system ' *Advances in Computer Aided Assessment,*

McDonald, A. (2002) 'The impact of individual differences on the equivalence of computer-based and paper-and-pencil educational assessments'. *Computers & Education*, 39 (3). pp 299-312.

McKenna, C. (2001) 'Academic approaches and attitudes towards CAA. a qualitative study', Danson, M and Eabry, C. eds). *5th International Computer Assisted Assessment Conference* Loughborough University Learning & Teaching Development, pp. 311-320

McKenna, C. & Bull, J. (2000) 'Quality assurance of computer-assisted assessment: practical and strategic issues'. *Quality Assurance in Education*, 8 (1). pp 24-32.

Mead, A & Drasgow, F. (1993) 'Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis' *Psychological Bulletin*, 114 pp 449-458.

Messick, S. (1989) 'Meaning and values in test validation: The science and ethics of assessment'. *Educational Researcher*, 18 (2). pp 5-12.

Miles, M. & Huberman, A. (1994) *Qualitative data analysis.* Thousand Oaks, CA Sage

Miller, F. & Varma, N. (1994) 'The effects of psychosocial factors on Indian children's attitudes toward computers'. *Journal of Educational Computing Research*, 10 pp 223-238

Mills, C (1999) *Development and introduction of a computer adaptive graduate records examinations general test.* eds. Drasgow, F. and Olsen- Buchanan, J.B , Innovations in computerized assessment. Lawrence Erlbaum.

Ministry of Education (2006) 'Educational Statistical year Book'. *Educational statistical Department., Sultanate of Oman*

Ministry of Health 'www.moh.gov om'. (Accessed: May 2010).

Ministry of Higher Education 'www mohe.gov.om,'. (Accessed· April 2010).

Ministry of Manpower 'www.manpower.gov.om'. (Accessed May 2010)

Ministry of Tourism (2010) 'Oman Geography'. (Accessed: May 2010 from www.omantourism.gov.om).

Morgan, D. (1988) 'Focus groups as qualitative research: Qualitative research methods series 16'.[in Newbury Park, CA. Sage. (Accessed:Morgan, D.

Morgan, D. (1997) *Focus groups as qualitative research.* Thousand oaks,CA Sage

Moroz, P. & Nash, J. (1997) 'Assessing and Improving the Factorial Structures of the Computer Self-Efficacy Scale', *Annual Meeting of the American Educational Research Association*. Chicago.

Murphy, P. (2001) 'Gendered learning and achievement'. *Understanding learning: Influences and outcomes*, pp 109-121.

Neil, P. M. (1996) 'The computerisation of the Army General Classification Test: A discussion paper.'. *Paper submitted for the 1996 E.F. Campbell medal.*

Neuman, G. & Baydoun, R. (1998) 'Computerization of paper-and-pencil tests: When are they equivalent?'. *Applied Psychological Measurement*, 22 (1). pp 71-83.

Nicol, D. (2007) 'E-assessment by design: using multiple-choice tests to good effect'. *Journal of Further and Higher Education*, 31 (1). pp 53-64.

NLLIA (1995) 'Cultural understanding as the eighth key competency.'. *Centre for Workplace Communication & Culture at University of Technology, Sydney*

O'Connell, T. & Dyment, J. (2006) 'Reflections on using journals in higher education: a focus group discussion with faculty'. *Assessment & evaluation in higher education*, 31 (6). pp 671-691.

O'Farrell, C. (2004) 'Enhancing Student Learning through Assessment. A Toolkit Approach'. *Centre for Academic Practice and Student Learning (CAPSL) Dublin Institute of Technology, 2004 Updated, 2009 University of Dublin Trinity College*, (Accessed: 2010 from www.cao/bd/pdf/250309_assessment_toolkit.pdf).

O'Malley, K., Kirkpatrick, R., Sherwood, W., Burdick, H., Hsieh, M. & Sanford, E. (2005) 'Comparability of a paper based and computer based reading test in early elementary grades', *The AERA Division D Graduate Student Seminar*. Montreal, Canada.

Oman Yearly Book (2009/10). (Accessed: 2010 from www.omanet.om/english/oman2009-2010/sec8_a.pdf.).

Pajares, F. (2001) 'Self-efficacy beliefs in academic settings'. *Review of Educational Research*, 66 (4). pp 543-578.

Pallant, J. (2005) *SPSS survival manual: A step by step guide to data analysis using SPSS version 12*. Maidenhead. UK: Open University Press

Palomba, C. & Banta, T. (1999) *Assessment Essentials: Planning, Implementing, and Improving Assessment in Higher Education. Higher and Adult Education Series.* Jossey-Bass, Inc., Publishers, Web site: http://www. josseybass. com

Papastergiou, M. (2010) 'Enhancing Physical Education and Sport Science students' self-efficacy and attitudes regarding Information and Communication Technologies through a computer literacy course'. *Computers & Education*, 54 pp 298-308.

Parshall, C. & Kromrey, J. (1993) 'Computer Testing versus Paper-and-Pencil Testing: An Analysis of Examinee Characteristics Associated with Mode Effect', *The Annual Meeting of The American Educational Research Association* Atlanta, GA

Peak, P. (2005) 'Recent trends in comparability studies'. *PEM research report*, pp. 05-05 (Accessed: 2007, from
http://www pearsonedmeasurement.com/downloads/research/RR_05_05.pdf)

Pearson (2009) 'Computer-Based & Paper- Pencil Test Comparability Studies'. *Test, Measurement & Research Services, Bulletin*,.(9), (Accessed: 2010.from
http://www pearsonassessments.com/NR/rdonlyres/93727FC9-96D3-4EA5-B807-5153EF17C431/0/Bulletin_9.pdf.).

Peterson, M., Augustine, C , Einarson, M & Vaughan, D. (1999) 'Designing student assessment to strengthen institutional performance in baccalaureate institutions'.[in Stanford, CA.: National Center for Postsecondary Improvement. (Accessed Peterson, M., Augustine, C., Einarson, M & Vaughan, D.

Poggio, J., Glasnapp, D., Yang, X. & Poggio, A. (2005) 'A Comparative Evaluation of Score Results from Computerized and Paper & Pencil Mathematics Testing in a Large Scale State Assessment Program'. *Journal of Technology, Learning, and Assessment*, 3 pp 1-30.

Pommerich, M (2004) 'Developing computerized versions of paper-and-pencil tests: Mode effects for passage-based tests'. *Journal of Technology, Learning, and Assessment*, 2 (6), pp. 1-30 (Accessed· 2007 from http·//escholarship.bc edu/jtla/vol2/6/).

Pommerich, M. & Burden, T. (2000) 'From Simulation to Application: Examinees React to Computerized Testing', *The Annual Meeting of the National Council on Measurement in Education*, . New Orleans, LA

Pomplun, M., Frey, S. & Becker, D. (2002) 'The score equivalence of paper-and-pencil and computerized versions of a speeded test of reading comprehension '. *Educational and Psychological Measurement*, 62 (2). pp 337-354.

Post-Kammer, P & Smith, P. (1985) 'Sex differences in career self-efficacy, consideration, and interests of eighth and ninth graders'. *Journal of Counseling Psychology*, 32 (4). pp 551-559.

Powell, R & Single, H. (1996) 'Focus groups'. *International journal for quality in health care: journal of the International Society for Quality in Health Care/ISQua*, 8 (5). pp 499-504.

QAA (2006) 'Code of practice for the assurance of academic quality and standards in higher education Section 6: Assessment of students ', (Accessed: 2009 from www.qaa.ac.uk/. ./CODEOFPRACTICE/SECTION6/COP_AOS.PDF).

Reinen, I. & Plomp, T. (1993) 'Some gender issues in educational computer use: Results of an international comparative survey'. *Computers & Education*, 20 (4). pp 353-365.

Richards, J., Platt, J & Weber, H. (1985) *Longman dictionary of applied linguistics*.

Richards, J & Schmidt, R (2002) *Longman Dictionary of Language Teaching and Applied Linguistics 3rd edition*. Essex,England:

Ricketts, C., Filmore, P., Lowry, R. & Wilks, S. (2003) 'How should we measure the costs of computer aided assessment?', *8th CAA Conference conference* Loughborough University, pp. 329-335.

Ricketts, C. & Wilks, S. (2001) 'Is Computer-Based Assessment Good for Students ?', Danson, M. and Eabry, C. eds.). *5th International Computer Assisted Assessment Conference* Loughborough University Learning & Teaching Development.

Ricketts, C. & Wilks, S. (2002a) 'What factors affect student opinions of computer-assisted assessment?', Danson, M. (ed. *6th International Computer Assisted Assessment Conference* Loughborough University Learning and Teaching Development, Loughborough University.

Ricketts, C. & Wilks, S. J. (2002b) 'Improving Student Performance Through Computer-based Assessment: insights from recent research'. *Assessment & evaluation in higher education*, 27 (5). pp 475 - 479.

Robson, C. (2002) *Real world research : a resource for social scientists and practitioner-researchers.* 2nd edn. Oxford, UK ; Madden, Mass.: Blackwell Publishers.

Rowntree, D. (1987) *Assessing students: How shall we know them?* London: 2nd edition,Kogan Page.

Russell, M. (1999) 'Testing Writing on Computers: A Follow-up Study Comparing Performance on Computer and on Paper'. *Educational Policy Analysis Archives*, 7 (20), (Accessed: 2007 from http://epaa.asu.edu/epaa/v7n20/).

Russell, M., Goldberg, A. & O 'Connor, K. (2003) 'Computer-based testing and validity: A look back into the future'. *Assessment in Education: principles, policy & practice*, 10 (3), pp. 279-293 (Accessed: 2007 from www.escholarship.bc.edu/cgi/viewcontent.cgi?article=1003&context=intasc).

Russell, M. & Haney, W. (1997) 'Testing Writing on Computers: An Experiment Comparing Student Performance on Tests Conducted via Computer and via Paper-and-Pencil'. *education policy analysis archives*, 5 (3), p. 3 (Accessed: 2007 from http://epaa.asu.edu/epaa/v5n3.html).

Sa'ada, J. & Saratawi, F. (2003) *Using Computers and Internet in Education Fields.* Jordan: A'Shurouq Press and Publishers

Sam, H., Othman, A. & Nordin, Z. (2005) 'Computer self-efficacy, computer anxiety, and attitudes toward the Internet: A study among undergraduates in Unimas'. *JOURNAL OF EDUCATIONAL TECHNOLOGYAND SOCIETY*, 8 (4), pp. 205-219 (Accessed: 2010 from www.citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.96.9844&rep).

Sawaki, Y. (2001) 'Comparability of Conventional and Computerized Tests of Reading in a Second Language'. *Language, Learning & Technology*, 5 (2), pp. 38-59 (Accessed: 2007 from http://.llt.msu.edu/vol5num2/sawaki/default.html).

Schaeffer, G., Reese, C., Steffen, M., McKinley, R. & Mills, C. (1993) 'Field test of a computer-based GRE general test (GRE Board Rep. No. 88-08P)'. *Princeton, NJ: Educational Testing Service,*

Schmidt, N., Norman, G. & Boshuzen, H. (1990) 'A cognitive perspective on medical expertise: theory and implications'. *Academic Medicine,* 65 pp 611-621.

Seale, C. (1999) 'Quality in qualitative research'. *Qualitative inquiry,* 5 (4). pp 465-478.

Seale, J. (2002) 'Using CAA to support student learning'. *LTSN Generic Centre Starter Guide,* 4 (Accessed: 2007 from www.alt.ac.uk/docs/eln004 pdf).

Sein, M., Olfman, L. & Bostrom, R. (1987) 'Training end users to compute: Cognitive, motivational and social issues.'. *INFOR,* 25 (3) pp 236-255.

Shashaani, L. (1997) 'Gender differences in computer attitudes and use among college students'. *Journal of Educational Computing Research,* 16 (1). pp 37-51.

Shashaani, L & Khalili, A (2001) 'Gender and computers: Similarities and differences in Iranian college students' attitudes toward computers'. *Computers & Education,* 37 (3). pp 363-375.

Shaw, K. (1996) 'Cultural issues in evaluation studies of Middle Eastern higher education'. *Assessment & evaluation in higher education,* 21 (4). pp 313-324.

Sheader, E., Gouldsborough, I & Grady, R. (2006) 'Staff and student perceptions of computer-assisted assessment for physiology practical classes' *Advances in Physiology Education,* 30 (4) pp 174-180.

Simpson, M. & Hayward, L (1998) 'Policy, research and classroom-based development: changing the assessment culture in Scottish schools'. *European Journal of Education,* pp 445-458.

Sly, L. & Rennie, L. (1999) 'Computer managed learning: Its use in formative as well as summative assessment', *3rd Annual CAA conference* Loughborough University Citeseer.

Smith, B. (2003) *Conventional versus computer based administration of measures of cognitive ability: an analysis of psychometric, behavioural, experiential and relativity of equivalence.* http://wwwlibrary.uow.edu.au/adtlNWU/public/adtlNWU20041006.142003/index.html. University of Wollongong

Smith, B. & Caputi, P. (2004) 'The development of the attitude towards computerized assessment scale'. *Journal of Educational Computing Research,* 31 (4). pp 407-422.

Smith, B., Caputi, P., Crittenden, N., Jayasuriya, R. & Rawstorne, P. (1999) 'A review of the construct of computer experience'. *Computers in Human Behavior,* 15 (2). pp 227-242

Space, L (1975) 'A console for the interactive on-line administration of psychological tests'. *Behavior Research Methods and Instrumentation,* 7 pp 191-193

Spray, J., Ackerman, T., Reckase, M. & Carlson, J. (1989) 'Effect of the medium of item presentation on examinee performance and item characteristics'. *Journal of Educational Measurement*, 26 (3). pp 261-271.

Staples, J. & Luzzo, D. (1999) *Measurement Comparability of Paper-and-pencil and Multimedia Vocational Assessments.* ACT.

Stephens, D. (1995) 'Results of a survey of the use of computer -assisted assessment in higher education in the UK', *Association of Learning Technology Conference.* University of Loughborough.

Stephens, D., Bull, J. & Wade, W. (1998) 'Computer-assisted Assessment: suggested guidelines for an institutional strategy'. *Assessment & evaluation in higher education*, 23 (3). pp 283-294.

Stephens, D. & Curtis, A. (2001) 'Use of Computer Assisted Assessment by Staff in the teaching of Information Science and Library Studies subjects'. *Innovations in Teaching And Learning in Information and Computer Sciences*, 1 (1), (Accessed: 2010 from http://www.ics.heacademy.ac.uk/italics/vol1iss1.htm).

Stephens, D. & Mascia, J. (1997) 'Results of a (1995) Survey into the use of Computer-Assisted Assessment in Institutions of Higher Education in the UK'. (Accessed: http://www.lboro.ac.uk/service/ltd/flicaa/downloads/survey/pdf).

Stephens, P. (2006) 'Validation of the business computer self-efficacy scale: Assessment of the computer literacy of incoming business students'. *Journal of Educational Computing Research*, 34 (1). pp 29-46.

Stephens, P. & Shotick, J. (2002) 'Re-evaluation of the computer self-efficacy model: development and use of the business computer self efficacy scale'. *Issues in Information Systems*, 3 pp 590-596.

Sternberg, R. (2007) 'Culture, instruction, and assessment'. *Comparative Education*, 43 (1). pp 5-22.

Stricker, L. & Wilder, G. (2001) 'Examinees' Attitudes About the TOEFL-CBT, Possible Determinants, and Relationships with Test Performance'. *RESEARCH REPORT-EDUCATIONAL TESTING SERVICE PRINCETON RR,*

Taylor, C., Jamieson, J., Eignor, D. & Kirsch, I. (1998) 'The relationship between computer familiarity and performance on computer-based TOEFL test tasks (TOEFL Research Report 61)'. *Princeton, NJ: Educational Testing Service,*

Texas Education Agency (2008) ' A Review of literature on the comparability of scores obtained from examinees on computer-based and paper-based tests'. *Technical report series,* (Accessed: 2010 from http://ritter.tea.state.tx.us/student.assessment/resources/techdigest/Technical_Reports/2008_literature_review_of_comparability_report.pdfTEXAS).

The Quality Assurance Agency for Higher Education (2007) 'Monitoring students' experiences of assessment: Enhancement Themes Guides to Integrative Assessment, no 1', (Accessed: available at www.enhancementthemes.ac.uk).

Thorpe, M., Edwards, R. & Hanson, A. (1993) *Culture and processes of adult learning: a reader.* Routledge

Torkzadeh, G., Koufteros, X. & Pflughoeft, K (2003) 'Confirmatory analysis of computer self-efficacy'. *Structural Equation Modeling: A Multidisciplinary Journal,* 10 (2) pp 263-275

Tsintsifas, A (2002) *A framework for the computer-based assessment of diagram-based coursework.* University of Nottingham.

Van Der Vleuten, C (1996) 'The assessment of professional competence: developments, research and practical implications'. *Advances in Health Sciences Education,* 1 (1). pp 41–67.

Van Der Vleuten, C. & Schuwirth, L. (2005) 'Assessing professional competence: from methods to programmes'. *Medical Education,* 39 (3). pp 309-317.

Wallace, A. (1999) *An exploratory study of the factors influencing the construction of computer self-efficacy.* Charles Sturt University.

Wallace, P. & Clariana, R. (2005) 'Gender difference in computer-Administered versions Paper-based Tests' *International Journal of Instructional Media,* 32 pp 37-45.

Wang, S (2004) 'Online or Paper: Does Delivery Affect Results?'. *Administration Mode Comparability study for Stanford Diagnostic Reading and Mathematics tests Harcourt assessment report.*

Wang, S , Jiao, H., Young, M , Brooks, T & Olson, J. (2007) 'A meta-analysis of testing mode effects in grade K-12 mathematics tests'. *Educational and Psychological Measurement,* 67 (2) pp 219.

Wang, S., Jiao, H., Young, M., Brooks, T. & Olson, J. (2008) 'Comparability of Computer-Based and Paper-and-Pencil Testing in K 12 Reading Assessments: A Meta-Analysis of Testing Mode Effects'. *Educational and Psychological Measurement,* 68 (1). pp 5-24.

Wang, S , Newman, L. & Witt, E (2000) 'AT&T aptitude test equivalence study: A comparison of computer and paper-and-pencil employment examinations. Bala Cynwyd'.[In PA: Harcourt Assessment System. (Accessed:Wang, S., Newman, L & Witt, E.

Wang, S., Young, M. & Brooks, T. (2004) 'SDRT 4/SDMT 4 Administration mode comparability study for Stanford Diagnostic Reading and Mathematics Tests'. (Accessed. 2007 from www.pearsonassessments com/NR/.../0/SDRT_SDMTCompStudy.pdf).

Waterfall, R (1979) 'Automating standard intelligence tests' *Journal of Visual Communication in Medicine,* 2 (1). pp 21-24.

Wiliam, D. (2008) 'What do you know when you know the test results? The meanings of educational assessments', *Annual Conference of the International Association for Educational Assessment* Cambridge, UK.

Winters, T. (2004) *Analysis, design, development, and deployment of a generalized framework for computer-aided assessment.* University of California.

Wolfe, E. & Manalo, J. (2004) 'Composition Medium Comparability in a Direct Writing Assessment of Non-Native English Speakers'. *Language, Learning & Technology*, 8 (1), (Accessed: 2009 http://llt.msu.edu/vol8num1/pdf/wolfe.pdf).

Wood, R. & Bandura, A. (1989) 'Impact of conceptions of ability on self-regulatory mechanisms and complex decision making'. *Journal of Personality and Social Psychology*, 56 (3), pp. 407-415 (Accessed: 2007 from www.des.emory.edu/mfp/Bandura1991JPSP.pdf -).

Wood, T. (2009) 'Assessment not only drives learning, it may also help learning'. *Medical Education*, 43 (1). pp 5-6.

www.enotes.com/topic/United_Kingdom.

www.mapsofworld.com.

www.omancensus.net.

www.statistics.gov.uk/statBase/ssdata.set.asp.

Zakrzewski, S. & Steven, C. (2000) 'A Model for Computer-based Assessment: the catherine wheel principle'. *Assessment & evaluation in higher education*, 25 (2). pp 201 - 215.

Zakrzewski, S. & Steven, C. (2003) 'Computer-based assessment: quality assurance issues, the hub of the wheel'. *Assessment & evaluation in higher education*, 28 (6). pp 609 - 623.

Zakrzewski, S., Steven, C. & Ricketts, C. (2009) 'Evaluating computer-based assessment in a risk-based model'. *Assessment & evaluation in higher education*, 34 (4). pp 439 - 454.

Zandvliet, D. & Farragher, P. (1997) 'A Comparison of Computer-Administered and Written Tests'. *Journal of research on computing in education*, 29 (4). pp 423-438.

Appendix A. Research Instruments

English Language Proficiency Test

Background information

Name(optional):				
Student academic number:				
Gender	(1) Male	(2) Female		
Region	(1) AlSharqiya	(2) AlBatinah	(3) AlDhahira	(4) AlDakhilia
College	(1) Sur	(2) Ibri	(3) Nizwa	
Do you have a computer at home?	(1) Yes	(0)No		

Part One: Grammar

Read the following statements and choose the correct answer.

1. Do you . .. . swim?
   a) can b) know how to c) are able to d) like

2. School was cancelled because it .... last night?
   a) rained b) rains c) raining d) rain

3. That book is ....
   a) fascinated b) fascinates c) fascinate d) fascinating

4. Salim and Abdulaziz .... come to work yesterday.
   a) don't b) didn't c) aren't d) do

5. Can you please give me .... piece of paper? This report is very long.
   a) other b) another c) these d) others

6. A: Did you talk to Ahmed last week?
B: No, I'll talk to .... tomorrow.

a) he   b) me   c) his   d) him

7. I live in a big city. .... many fun places to visit.
a) It have   b) They are   c) There are   d) They have

8. I .... on my way to class.
a) go   b) am going   c) am   d) do

9. Mexico City isn't the capital of Mexico, is it?
a) Yes, it does   b) Yes, it isn't   c) Yes, it is   d) No, it is

10. Fatima is cooking lunch .... the kitchen.
a) by   b) next to   c) at   d) in

11. Brazil and Peru .... South American countries.
a) is   b) are   c) have   d) were

12. .... book next to me is interesting.
a) These   b) Those   c) This   d) That

13. I'm going to .... store. Do you need anything?
a) my   b) some   c) the   d) a

14. Do you like .... car?
a) Michael   b) Michael's   c) Michaels   d) Michaels'

15. I was Surprised because the cat .... the ball easily.
a) caught   b) catched   c) catches   d) will catch

16. What . ... thinking about now?
a) are you b) do you c) is d) you are

17. How .... spell "happy" ?
a) you do b) do you c) is you d) you are

18. I decided to . ... to North Africa last summer.
a) travelling b) travelled c) travel d) traveler

19. Young people need a lot of .... to choose a job.
a) advise b) advice c) advises d) advices

20. I ... . been a good singer.
a) was b) has not c) have never d) have ever

21. Before last summer, I ...... never flown on an airplane.
a) have b) had c) did d) was

22 When did you ....?
a) married b) got married c) get married d) get marriage

23. Please return the video ...... borrowed from me last month.
a) that it b) where you c) that you d) that you have

24. I need . . ... information about the wedding party.
a) one b) much c) some d) a few

25. A: . ... . that movie fantastic?
B: Yes, it definitely was.

a) Was b) Wasn't c) Did d) Were

26. A: Let's go on a picnic.
B: OK, when do you want to have .... picnic?

a) a		b) an		c) that		d) the

27. Although .... late for class, the teacher let him in.
a) he is		b) he comes		c) he is coming		d) he was

28. The boy .... a picture seems to be a tourist.
a) taken		b) taking		c) who was taken		d) takes

29. He .... scuba diving in the ocean.
a) doesn't want		b) wants to		c) hopes		d) didn't like

30. A: Do you want to eat out tonight?
B: No, I ...... cook at home.

a) would rather		b) prefer		c) need		d) shouldn't

Part Two: Reading Comprehension

Text One:

The Old Man and His Grandson

There was once a very old man, whose eyes had become dim, his ears dull of hearing, his knees trembled, and when he sat at table he could hardly hold the spoon, and spilt the broth upon the table-cloth or let it run out of his mouth. His son and his son's wife were disgusted at this, so the old grandfather at last had to sit in the corner behind the stove, and they gave him his food in an earthenware bowl, and not even enough of it. And he used to look towards the table with his eyes full of tears.

Once, too, his trembling hands could not hold the bowl, and it fell to the ground and broke. The young wife scolded him, but he said nothing and only sighed. Then they brought him a wooden bowl for a few half-pence, out of which he had to eat.

They were once sitting thus when the little grandson of four years old began to gather together some bits of wood upon the ground. 'What are you doing there?' asked the father. 'I am making a little trough,' answered the child, 'for father and mother to eat out of when I am big'.

The man and his wife looked at each other for a while, and presently began to cry. Then they took the old grandfather to the table, and henceforth always let him eat with them, and likewise said nothing if he did spill a little of anything

Choose the correct answer to these questions. Each question has only one correct answer

1) Which health problem was NOT mentioned in the fairy tale?

a) Poor vision
b) Bad hearing
c) Terrible headaches

2) Which action caused the son to move the old grandfather to sit in the corner behind the stove?

a) He dropped his cane
b) He spilt his broth.
c) He didn't pay attention

3) Which action caused them to exchange a wooden bowl for the earthenware bowl?

a) He broke his bowl
b) He complained about the taste
c) He insulted his grandson

4) What action caused them to reconsider their treatment of the old man?

a) The old man looked towards the table with his eyes full of tears.
b) Their child began creating something for them to eat out of in their old age
c) Their child began crying when speaking with the old man.

Text Two:

The Importance of English

British or American, the language is basically the same, and its global stature is backed up by massive English-language training programmes, an international business that in textbooks, language courses, tape cassettes, video programmes and computerized instruction — is worth hundreds of millions of pounds or dollars to the economies of the US and the UK. The English language is now one of Britain's most reliable exports.

In the ironic words of the novelist Malcolm Bradbury, it is an ideal British product, 'needing no workers and no work, no assembly lines and no assembly, no spare parts and very little servicing, it is used for the most intimate and the most public services everywhere. We call it the English language ...'Dr Robert Burchfield, former Chief Editor of the Oxford English Dictionary, has remarked that 'any literate, educated person on the face of the globe is deprived if he does not know English'. The first level of the global sway of English is to be found in those countries, formerly British colonies, in which English as a second language has become accepted as a fact of cultural life that cannot be wished away. In Nigeria, it is an official language; in Zambia, it is recognized as one of the state languages; in Singapore, it is the major language of government, the legal system and education.

1. Are the following statements true or false?

a. The teaching of the English language worth a lot of money. ___

b. Dr. R. Burchfield is not the Chief Editor of the Oxford English Dictionary currently. ___

c. Malcolm Bradbury is a novelist that works at an assembly line. ___

2. Find a word or phrase in the text which, in context, is similar in meaning to:

a. Essentially: ____________________

b. That can be trusted: ____________________

3. Choose a, b or c in each question below. Only one choice is correct.

1 The English Language

a. is a good source of money for UK and USA.

b. is not a good source of money for UK and USA.

c. is a good source.

2. The English Language

a. has not been accepted in the British colonies.

b. has been accepted in the British colonies.

c. has become obsolete in the British colonies.

3. The English language is an official language

a in Zambia and Singapore

b. in Nigeria.

c. in Nigeria and Zambia.

Text Three:

The Titanic: Another Disaster Movie?

The biggest gamble in movie making history has received no fewer than fourteen Oscar nominations. And not only that· contrary to what almost everyone said during the troubled course of its production, it looks like the gamble is going to pay off. Titanic, James Cameron's $200 million epic has been a great success across Europe and America. Test screenings in America have been overwhelmingly positive. Daily Variety, the most influential film newspaper in America, has no doubts. They call it a spectacular demonstration of what modern technology can contribute to dramatic story-telling, and *concludes that, unlike the liner which sank in the North Atlantic eighty-five years* ago, this Titanic arrives at its destination.

But it hasn't been all that easy. Shot over eight months (about two months longer than originally planned), Titanic saw its budget spiral almost out of control. With such lavish sets, huge casts and expensive, state-of-the-art computerized special effects, the movie seemed set to become as big a disaster as the story on which it was based.

The perfectionist James Cameron went to extraordinary lengths to get his film exactly right. The filmmakers built a perfect scale model of the ship, which at 236 meters was almost as big as the original, and mounted it on giant hydraulic lifts to reproduce the effect of the vessel sinking The model stood in a tank filled with 77million litres of sea water. Everything in Cameron's movie is bigger, grander and more expensive than anything that has been done before.

1. Read each definition below and choose the word from the list that matches the definition.

gamble	success	budget	vessel

a A large ship or boat·

b. A risk taken in order to gain some advantage:

c. The cost or estimated cost of something:

d. A favourable, satisfactory result or accomplishment:

2. Complete the following paragraph with the correct form of the verb in brackets.

Negative publicity (**begin**) (a)____________ early on in production, when 50 crew members, including Cameron himself, (**eat**) (b)_______________ a meal of lobster soup which (**poison**) (c) _______________ with the drug PCP, possibly by a discontented worker. Then other stories about terrible working conditions (**start**)

(d) ________________ to make the headlines.

END OF EXAMINATION

Computer Experience and Computer Self-Efficacy Questionnaire

Background Information

Name (optional)				
Student academic number:				
Gender	(1) Male	(2) Female		
Region	(1) AlSharqiya	(2) AlBatinah	(3) AlDhahira	(4) AlDakhilia
College	(1) Sur	(2) Ibri	(3) Nizwa	
Do you have a computer at home?	(1) Yes	(0)No		

Dear students:

I will be very appreciative if you complete this questionnaire. The questionnaire consists of two sections. Section one related to the computer experience which demonstrate the amount and kind of computer use. Section two related to the computer self efficacy winch demonstrate how confidents you are in computer skills. Your answer will give a valuable feedback.

Section One (8 items):

(1) Which choice below best describes your computer experience OR

How do you rate your overall computer experience?

1-Real computer pro 2- lots of experience 3-moderate experience

(2) How often do you use computer at home?

1- Most days 2- About 2-3 days a week 3- less than one per a week

(3) How often do you use computer at university?

1- Most days 2- About 2-3 days a week 3- less than one per a week

(4) How many years ago did you first begin using computers?

1- More than five years 2- between 2 to 4 years 3- less than one year

(5) How is your knowledge about computers and its softwares?

1- More than adequate for my work 2-Adequate for my work 3- Less than adequate for my work

(6) Do you use computers to send emails?

1- Often 2- Some times 3- Rarely

(7) Do you use computers to play games?

1- Often 2- Some times 3- Rarely

(8) Do you use computers to print assays, letter, etc?

1-Often 2- Some times 3- Rarely

Section two (7 items):

Please answer the following questions according to your feelings of confidence for successfully performing the specified task.

1. Very Little Confidence. 2. Little Confidence. 3. Some Confidence. 4. Moderate Confidence.

5. High of Confidence

	Statements	Rate of Confidence
1.	I feel confident opening a data file to view on the monitor screen.	(Low) (High) 1 2 3 4 5
2.	I feel confident using the computer to write an essay or a letter.	1 2 3 4 5
3.	I feel confident entering and saving data (numbers or words) into a file.	1 2 3 4 5
4.	I feel confident making selections from an on-screen menu.	1 2 3 4 5
5.	I feel confident escaping/exiting from a program or software.	1 2 3 4 5
6.	I feel confident working on a personal computer (micro computer).	1 2 3 4 5
7.	I feel confident using a printer to make a "hardcopy" of my work.	1 2 3 4 5

Thank you for your time and cooperation

Adapted from:

Fagan, M.H, Neill, S and Wooldridge, B.R. 2004. An empirical investigation into the relationship between computer self-efficacy, Anxiety, Experience, Support and usage. Journal of Computer information Systems. Winter 2003-2004 pp 95-104.

Johnson, D. M., Ferguson, J. A., and Lester, M.L, 1999. Computer experience, self efficacy and Knowledge of students enrolled in introductory university agriculture courses. Journal of Agricultural Education. 40(2) 28-37.

Attitude Towards Computerized Assessment Scale (ATCAS)

Background Information

Name(optional)				
Student academic number:				
Gender	(1) Male	(2) Female		
Region	(1) AlSharqiya	(2) AlBatinah	(3) AlDhahira	(4) AlDakhilia
College	(1) Sur	(2) Ibri	(3) Nizwa	
Do you have a computer at home?	(1) Yes	(0)No		

Dear student:

Listed below are a series of statements describing various thoughts and feelings, which you may have about completing *tests on a computer*. Please indicate whether you agree or disagree with each statement by circling the most appropriate response. Use the following scale to guide your response to each statement:

1= *Strongly Disagree*

2= Disagree

3= Uncertain

4= Agree

5= Strongly Agree

1. I felt more nervous completing the test on a computer than on a paper.
   (Strongly Disagree) 1 2 3 4 5 (Strongly Agree)

2. The test instructions presented on the computer were difficult to understand.

(Strongly Disagree) 1 2 3 4 5 (Strongly Agree)

3. I would have found it helpful if I was given more practice time on the computer before starting the test.

(Strongly Disagree) 1 2 3 4 5 (Strongly Agree)

4. Reading an item/question on the computer screen was more difficult than reading the same item from the paper-and-pencil test form

(Strongly Disagree) 1 2 3 4 5 (Strongly Agree)

5. Having to answer item/questions by using the computer keyboard was easier than handling a separate response sheet.

(Strongly Disagree) 1 2 3 4 5 (Strongly Agree)

6. I felt more anxious taking the test on the computer than on paper.

(Strongly Disagree) 1 2 3 4 5 (Strongly Agree)

7. Worrying about my lack of computer experience interfered with my performance on the computer administered test.

(Strongly Disagree) 1 2 3 4 5 (Strongly Agree)

8. I would rather take a computer-administered test than a paper-and-pencil test in the future.

(Strongly Disagree) 1 2 3 4 5 (Strongly Agree)

9. Computerized tests require too much experience with computers.

(Strongly Disagree) 1 2 3 4 5 (Strongly Agree)

10. I wish computerized tests did not bother me so much.

(Strongly Disagree) 1 2 3 4 5 (Strongly Agree)

11. Thinking about pressing the wrong key interfered with my performance on the computer administered test.

(Strongly Disagree) 1 2 3 4 5 (Strongly Agree)

12. It was easier to check my responses on the paper-and-pencil test than on the computer administered test.

(Strongly Disagree) 1 2 3 4 5 (Strongly Agree)

13. I felt more comfortable completing the test on paper than on the computer.

(Strongly Disagree) 1 2 3 4 5 (Strongly Agree)

Thank you for your time and cooperation.

Adapted from: Smith, B and Caputi, P (2004). The development of the attitude towards computerized assessment scale. Educational Computing Research 31(4): 407-422.

Staff Questionnaire

Gender:	(1) Male	(2) Female	
Nationality/Language:	(1) Omani	(2) Arabic/ Non-Omani	(3) English/Non-Omani

Dear:

I will be very appreciative if you complete this questionnaire. Your answer will give a valuable feedback to know your opinion towards Computer Assisted Assessment (CAA).

Evaluate the following statements by writing the number which expresses your agreement:

1 = Strongly disagree.

2 = Disagree.

3 = Moderate.

4 = Agree.

5 = Strongly agree.

<u>THANK YOU FOR YOUR COOPERATION AND TIME.</u>

1 = Strongly disagree. 2 = Disagree. 3 = Moderate. 4 = Agree. 5 = Strongly agree.

	Statements	Agree ment Score
1	I think computer is very useful.	
2	I like more exams to be administrated by using Computer Assisted Assessment (CAA).	
3	I use computer at my work.	
4	I think computer test is easier than paper test.	
5	I feel confident delivering tests by computers.	
6	I think questions will be very clear in the screen.	
7	I think the instructions will be easy in computer.	
8	I use computer in my house.	
9	I think implementing CAA will face lots of barriers.	
10	I dislike assessing student by machines.	
11	I think computerized test can only assess low cognitive ability.	
12	I hesitate to use computer for fear of losing all my work.	
13	I think using CAA in the second language is difficult.	
14	I think computerized tests take longer to complete then paper test.	
15	I feel comfortable using CAA with my students.	
16	I like to attend computerized test training.	
17	I prefer using computerized test than paper test.	
18	I think CAA will save my time.	

STUDENT FOCUS GROUP

Number of Male:	**Number of Female:**
Date:	**College:**

Questions Before and After Taking the Test

1- If you are given a choice between sitting a test on Computer or using paper and pencil; what will be your choice? Why?

2- Does/Did anything worry you about computer assessment? What? Why

3- What is the thing you will like /did like most about taking a test on a computer?

4- What is the thing you disliked most about taking a test on a computer?

5- From your point of view, which part will be /was easier in computerized test the Grammar or Reading? Why?

6- Will you/Did you find reading passages in the computerized test easy or difficult? Why?

7- Do you think that there will be any differences in taking CAA in your mother tongue or in English? If Yes or No Why?

8- Which method will you/ did you prefer for testing- computer or paper and? Why?

Questions After Taking the Test

9- Using one word. How you will express your overall experience being assessed by using computers? Why?

10- Did you face any difficulties during the exam by computer?

11- Explain. How do you feel before and after having the CAA?

Thank you for your cooperation and time

Staff Semi-Structured Interview

Gender:	Male	Female	
Nationality:	Omani	Arabic/Non- Omani	English/Non -Omani
Date:			

Part One Questions

1- Do you have computer at home?

2- How would you describe your computer experience?

3- Do you use computer in your academic work? If Yes what kind of work?

4- Do you know about Computer Assisted Assessment? If No, I give brief information about CAA.

5- If Yes. Did you use Computer Assisted Assessment?

Part Two Questions

1- From your point of view, what are the difficulties that some institutes might face if they intend to implement CAA?

2- What do you think the advantages in using Computer Assisted Assessment are? Can you explain that?

3- What do you think the disadvantages in using Computer Assisted Assessment are? Can you explain that?

4- What are the limitations you think you will encounter when using the CAA?

5- Do you have any suggestions for moving forward towards CAA?

Part Three Questions

6- What question types do you think that you can use in CAA? Why?

7- Which method of assessment do you think will longer to complete: paper and pencil test or computerized test? What are the reasons?

8- If the institute has software or a programme about CAA, will you use it in the formative and summative assessment? If No or Yes Why?

9- Computerized tests require training the academic staff, if the institute presented this training will you participate in it?

10-*How would you feel about switching from current assessment (paper-*pencil test) to computerized test?

11-If that happen, what is the big concern you will have towards switching?

12-Overall, how do you rate the CAA compared with paper-pencil assessment?

13-Is there anything else in this regard you wish to mention?

Thank You for Your valuable Time.

Student Participant Information Sheet

Computer Assisted Assessment in Oman: factors affecting student performance.

Dear student, you are invited to participate in this research project, but before you decide to participate it is important for you to know the aims of this research and what is required from you. So, please read the following information carefully and feel free to ask if there is anything that you want to know or any further inquiry that you need explained. Finally it is for you to decide to participate or not.

What is the aim of this research?

The aim of this research is to investigate and explore the feasibility of computer aided assessment use in the Sultanate of Oman. This research holds three main dimensions and they are:

First dimension: comparing students' performance between paper-pencil assessment and computerised assessment.

Second dimension: investigating the attitudes of both the students and the academic staff towards computerised assessment.

Third dimension: exploring students' and staff thoughts and feelings in depth towards computer assisted assessment as well as the difficulties and hurdles which might face CAA from their points of view when applying it.

What kind of participants does this study require?

This study aims to investigate whether there is any difference in performance between paper-based and computer-based assessments in the foundation year and, if so, whether it is affected by gender, college, region or computer experience. We will be sampling from four regions (Sharqiya, A'Dhahira, AlBatina and AlDakhiliya) and three colleges (Sur, Ibri and Nizwa) and that is why you have been invited to take part in this study

What is the role of the participants in this study?

The participants will be asked to take a placement test in English language in either a computerized or paper-based version, but before they sit for the test a questionnaire will be distributed to them to evaluate the level of their computer experience. After taking the computerized test, participants will be asked to answer an online questionnaire about their attitudes towards computer aided assessment. Your results will be revealed to you only if you want and your personal details will remain confidential.

Some participants will be asked to participate in group interviews to investigate their feeling, views and thoughts towards computer aided assessment in depth. These group interviews may take place both before and after the English language test.

Can the participant change his/her mind and withdraw him/herself from participating in the study?

You have total freedom to withdraw yourself from the study at any time you wish without giving any reasons for your decision and without any further consequences.

Are there any disadvantages if I participated in the study?
There are no disadvantages except you may be disappointed in your performance on the test. This study was approved by the University of Plymouth in UK which assured that this research process is compatible with the ethics of research. Furthermore, this study was also approved by the Ministry of Higher Education in the Sultanate of Oman.

What are the possible advantages if I participated in this study?
This study will provide you with useful information about your level in the English test which will help you to take decisions to come out with suitable learning plan. In addition, your participation will help you understand your attitudes, as well as your colleagues' attitudes, towards computer aided assessment.

Will my participation in this study and the collected data will be confidential?
In this study all the data which include your name, academic number, gender, region, college and the results as previously mentioned will be confidential and will be kept in a safe place where the researcher only could access them.

What will happen to the study's results?
This study is research towards a PhD degree. The results will be used only for this purpose and will be presented in the PhD dissertation. They may also be published in a scientific journal or scientific conference In all forms of publication you will not be identified personally, any results and remarks will be anonymised and you can ask for a copy of the results if you want

Who is funding this research?
This research is funded by the Ministry of Higher Education in the Sultanate of Oman

What if I faced any problem?
If you have any inquiry related to any part of the study you are welcome to ask the researcher who will help and answer all your questions. You are also welcome to make your views and suggestions know to the researcher.

If you need any further information you can contact:
amina.al-hajri@plymouth ac uk am.alhajri@yahoo.com
Telephone: 0096895226825 or 00447726245668
Or you can write to the following address:
Po.Box.170, PC. 421. Bidiya, Sharqiya North, Sultanate of Oman.

I appreciate your time spent in reading the above information and considering whether to participate in this study. If you have decided to participate in this study please sign the attached consent form.

Faculty of Technology

Computer Assisted Assessment in Oman: factors affecting student performance

Amina Al-Hajri amina.al-hajri@plymouth.ac.uk

Student Consent form for participants

I have read all the attached information about the study and its aims and understood everything related to it. I have also found all the answers to the questions I wanted to ask and I am satisfied with them. I have also understood that I have the freedom to ask for any further information at any time along with the following:-

1- My participation in this study is totally voluntary.
2- I have the freedom to withdraw at any time without any consequences.
3- All the information I will provide will be kept in safe place and confidential.
4- All the data related to the study could be published in either scientific journal or conference, but my identity will be kept anonymous.
5- I agree to the focus group being audio recorded.
6- I agree to the use of anonymised quotes in publications

Therefore, I agree to participate in this research.

Name of the participant: Researcher's Name:

Signature: Signature:

Date: Date

Staff Participant Information Sheet

Computer Assisted Assessment in Oman: factors affecting student performance.

Dear colleague, you are invited to participate in this research project, but before you decide to participate it is important for you to know the aims of this research and what is required from you. So, please read the following information carefully and feel free to ask if there is anything that you want to know or any further inquiry that you need explained. Finally it is for you to decide to participate or not.

What is the aim of this research?
The aim of this research is to investigate and explore the feasibility of computer aided assessment use in the Sultanate of Oman. This research holds three main dimensions and they are.
First dimension: comparing students' performance between paper-pencil assessment and computerised assessment.
Second dimension: investigating the attitudes of both the students and the academic staff towards computerised assessment
Third dimension: exploring students' and staff thoughts and feelings in depth towards computer assisted assessment as well as the difficulties and hurdles which might face CAA from their points of view when applying it

What kind of participants does this study require?
This part of the study aims to investigate the attitudes academic staff (Omani and non-Omani) towards computer aided assessment. We will be sampling in three of the Applied Colleges in Oman (Sur, Ibri and Nizwa): that is why you have been invited to take part in this study.

What is the role of the participants in this study?
The participants will be asked to answer a questionnaire to explore their attitudes towards computer aided assessment. In addition, some participants will be interviewed individually to investigate their thoughts and feelings towards computerized assessment in greater depth.

Can the participant change his/her mind and withdraw him/herself from participating in the study?
You have total freedom to withdraw yourself from the study at any time you wish *without giving any reasons for your decision and without any further* consequences.

Are there any disadvantages if I participated in the study?
There are no disadvantages that we are aware of. This study was approved by the University of Plymouth in UK which assured that this research process is

compatible with the ethics of research. Furthermore, this study was also approved by the Ministry of Higher Education in the Sultanate of Oman.

What are the possible advantages if I participated in this study?
Your participation will help you understand your attitudes, as well as your colleagues' attitudes, towards computer aided assessment.

Will my participation in this study and the collected data will be confidential?
In this study all the data which include your name, gender, or college will be confidential and will be kept in a safe place where the researcher only could access them.

What will happen to the study's results?
This study is research towards a PhD degree. The results will be used only for this purpose and will be presented in the PhD dissertation. They may also be published in a scientific journal or scientific conference. In all forms of publication you will not be identified personally, any results and remarks will be anonymised and you can ask for a copy of the results if you want.

Who is funding this research?
This research is funded by the Ministry of Higher Education in the Sultanate of Oman.

What if I faced any problem?
If you have any inquiry related to any part of the study you are welcome to ask the researcher who will help and answer all your questions. You are also welcome to make your views and suggestions know to the researcher.
If you need any further information you can contact:
amina.al-hajri@plymouth.ac.uk am.alhajri@yahoo.com
Telephone: 0096895226825 or 00447726245668
Or you can write to the following address:
Po.Box:170, PC: 421, Bidiya, Sharqiya North, Sultanate of Oman.

I appreciate your time spent in reading the above information and considering whether to participate in this study. If you have decided to participate in this study please sign the attached consent form.

Faculty of Technology

Computer Assisted Assessment in Oman: factors affecting student performance

Amina Al-Hajri amina.al-hajri@plymouth.ac uk

Staff Consent form for participants

I have read all the attached information about the study and its aims and understood everything related to it. I have also found all the answers to the questions I wanted to ask and I am satisfied with them. I have also understood that I have the freedom to ask for any further information at any time along with the following:-

1-My participation in this study is totally voluntary.

2-I have the freedom to withdraw at any time without any consequences.

3-All the information I will provide will be kept in safe place and confidential.

4-All the data related to the study could be published in either scientific journal or conference, but my identity will be kept anonymous.

5-I agree to the interview being audio recorded.

6-I agree to the use of anonymised quotes in publications.

Therefore, I agree to participate in this research.

Name of the participant: Researcher's Name:

Signature: Signature:

Date: Date.

Appendix B. Students Performance on Paper and Computer Mode

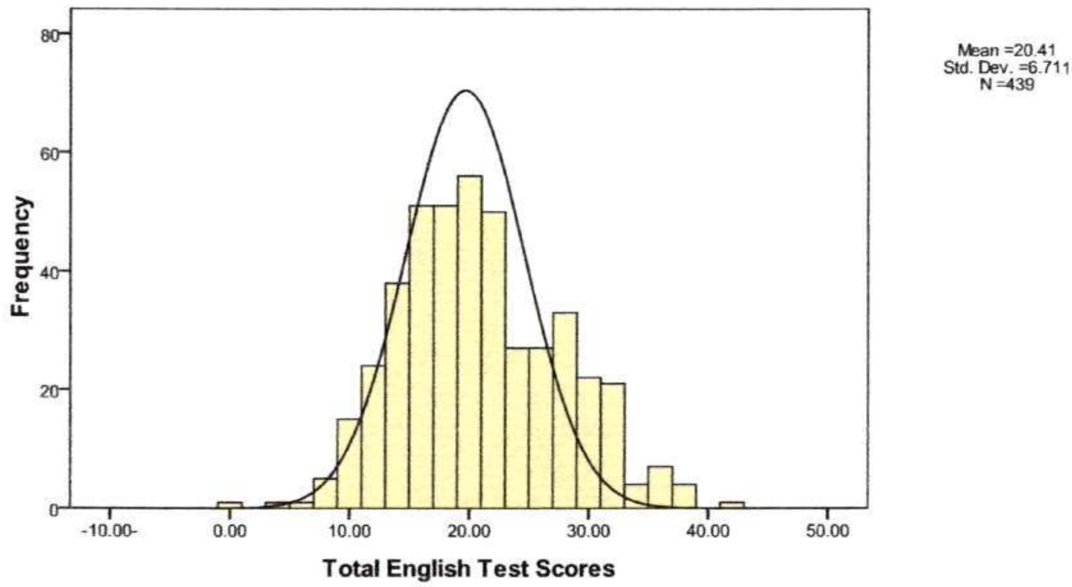


Figure 1. Histograms for the Total English Test Scores

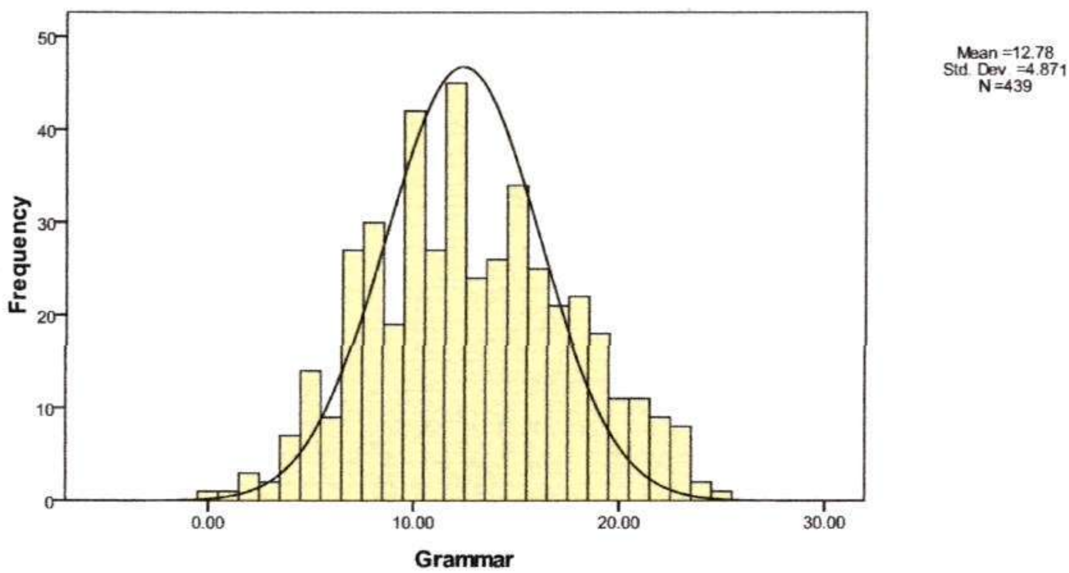


Figure 2. Histograms for Grammar

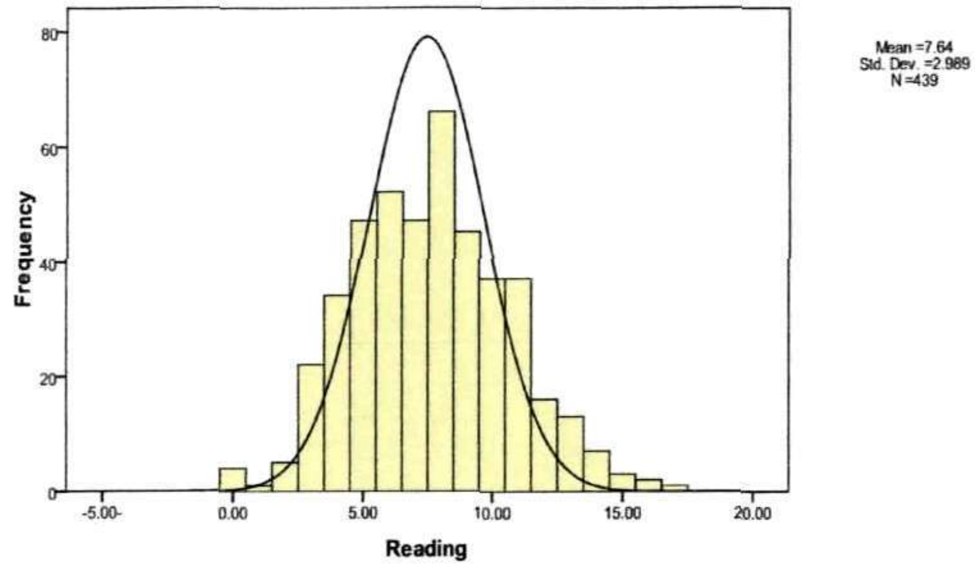


Figure 3. Histograms for Reading

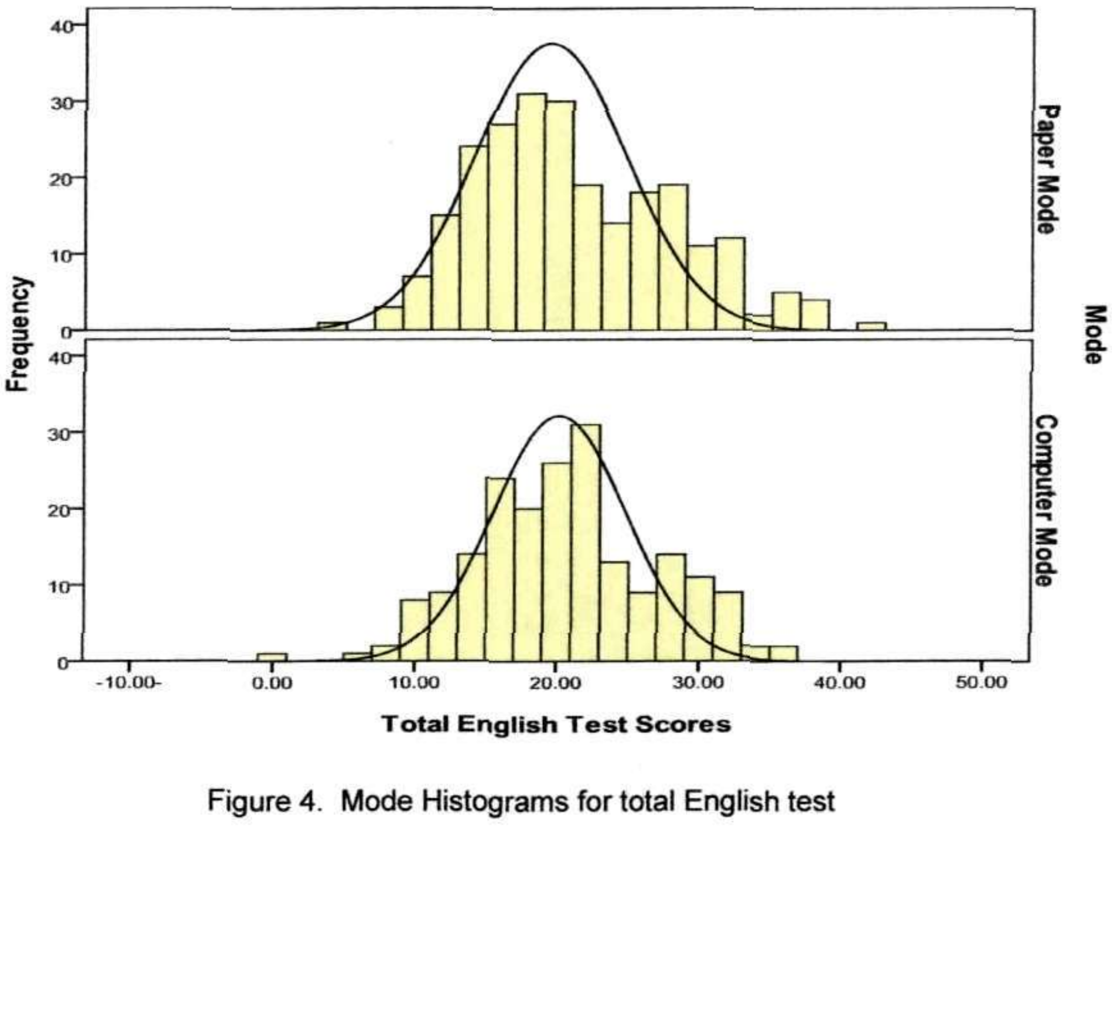


Figure 4. Mode Histograms for total English test

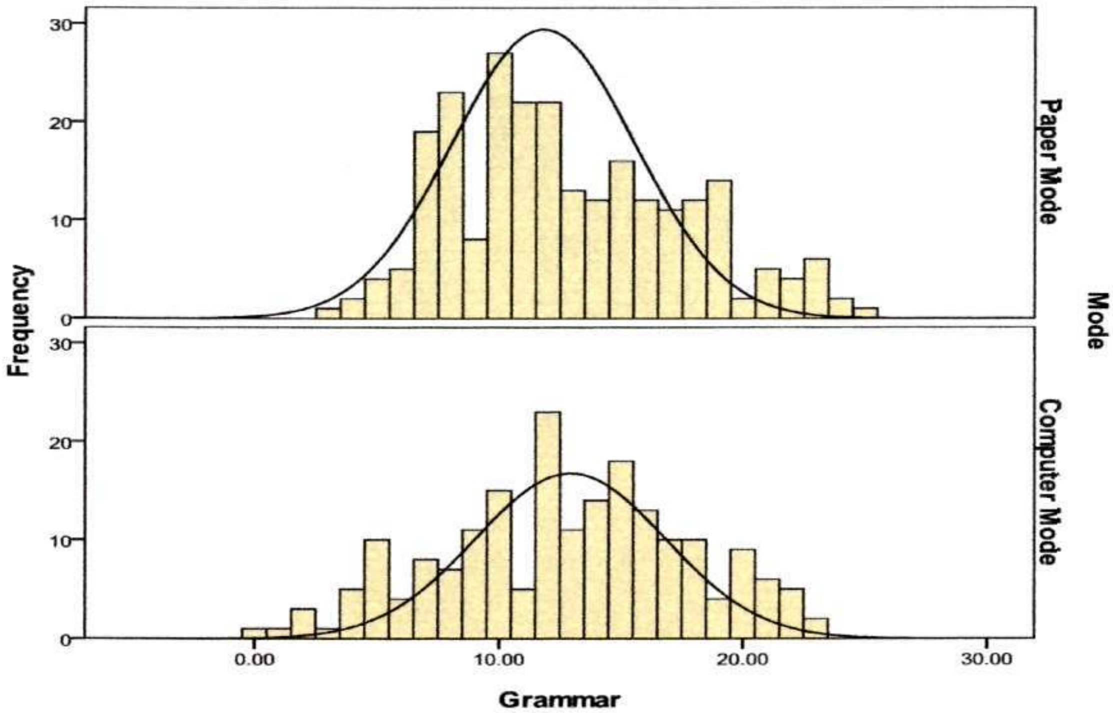


Figure 5. Mode Histograms for Grammar

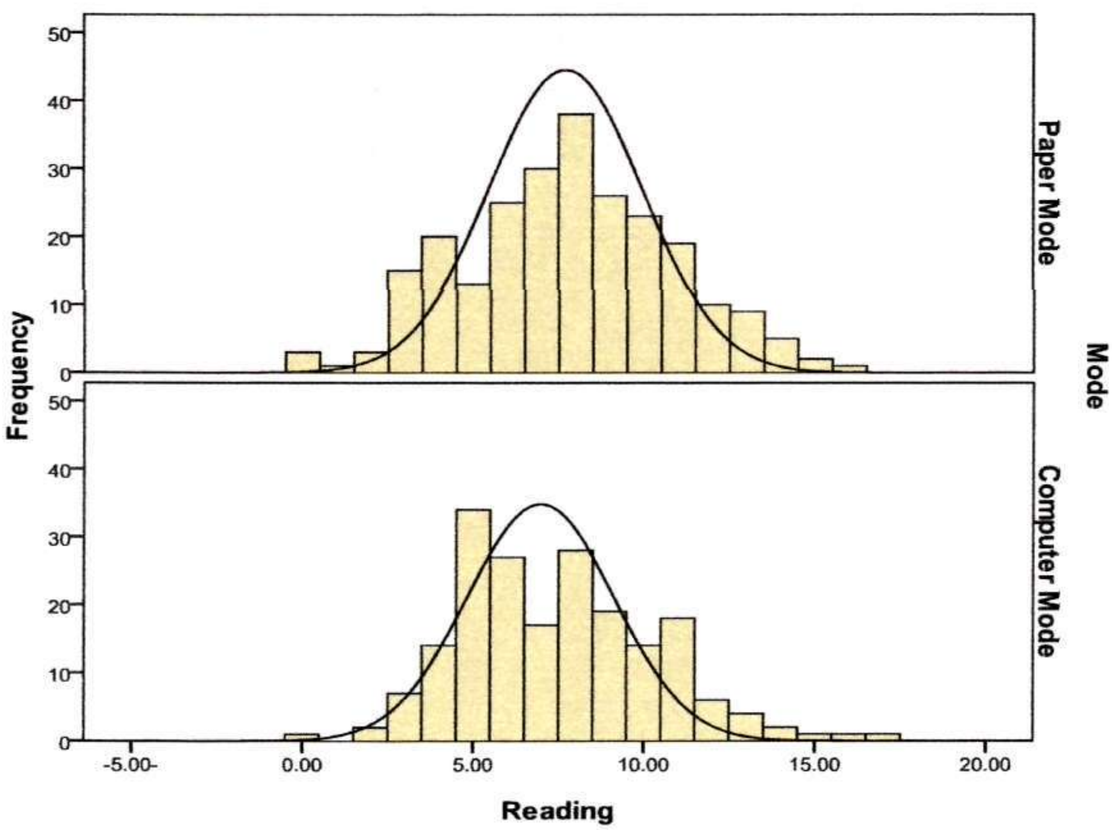


Figure 6. Mode Histograms for Reading

Mode	*N*	*M*	*SD*	*T-test*	*Sig(2-tailed)*
Paper	*243*	*20.55*	*6.95*	*.504*	*.615*
Computer	*196*	*20.23*	*6.41*		

Table 1. Means, Standard Deviations and T. Test Result for the Effect of Mode on Total English Test

	Mode	*N*	*M*	*SD*	*T-test*	*Sig(2-tailed)*
Grammar	*Paper*	*243*	*12.79*	*4.71*	*.073*	*.942*
	Computer	*196*	*12.67*	*5.06*		
Reading	*Paper*	*243*	*7.76*	*3.078*	*1.014*	*.311*
	Computer	*196*	*7.47*	*2.87*		

Table 2. Means, Standard Deviations and T. Test Result for the Effect of Mode on Grammar and Reading Test

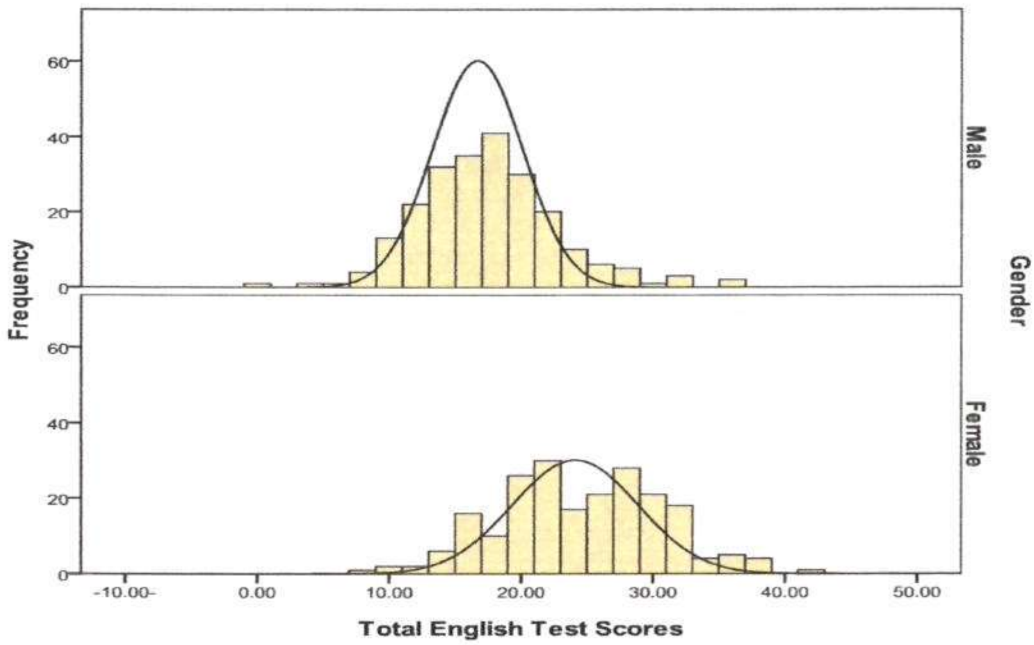


Figure 7 Gender Histograms for the Total English Test Scores

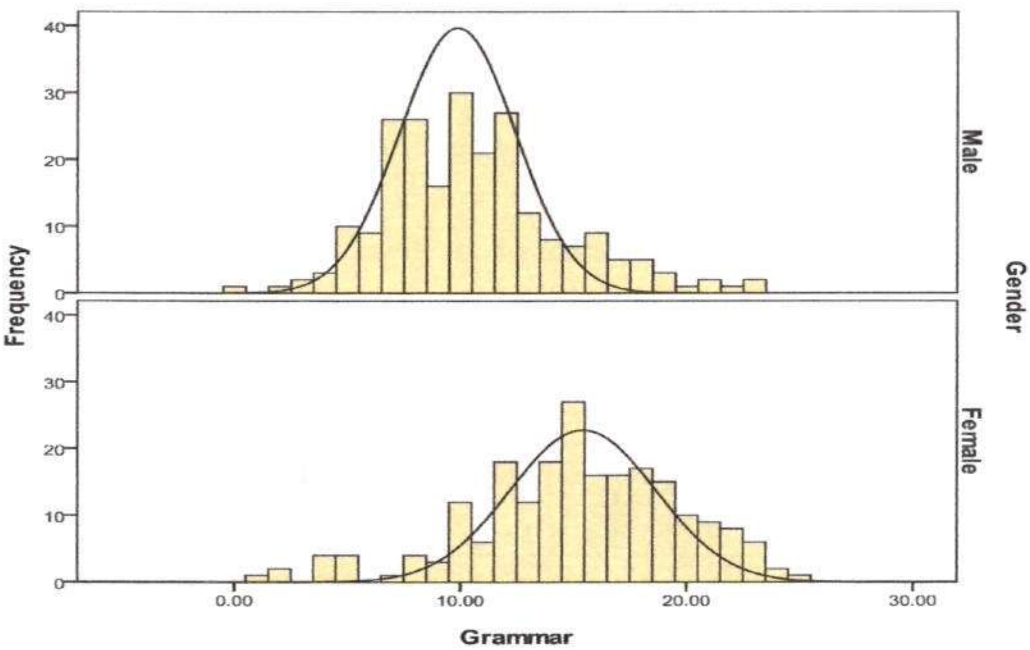


Figure 8. Gender Histograms for Grammar

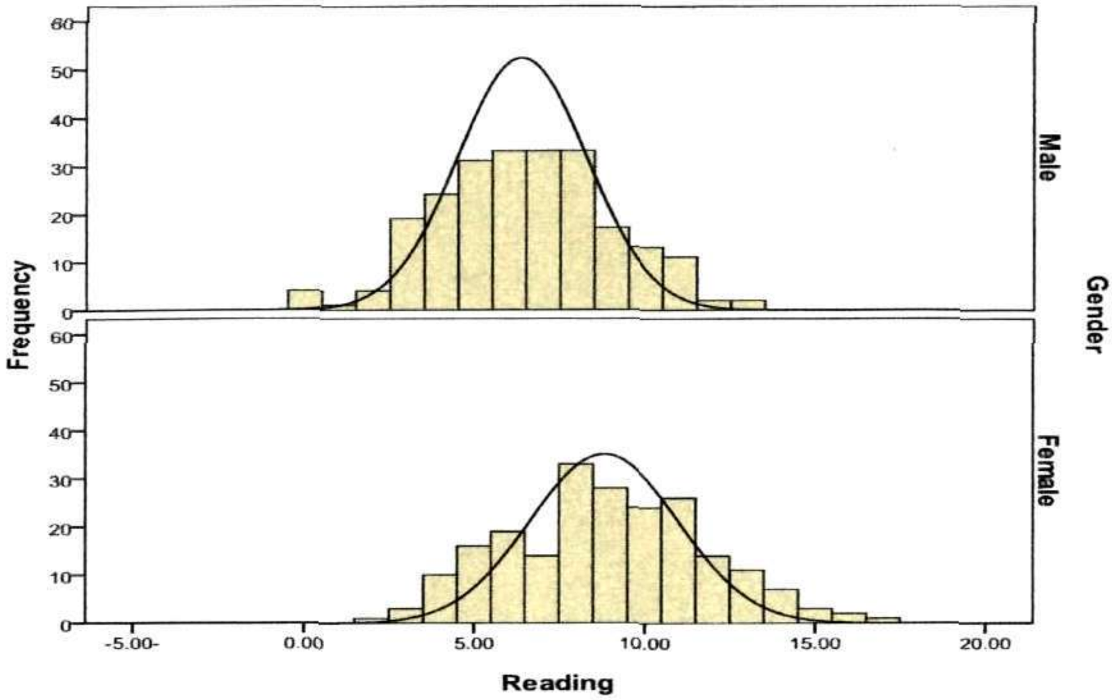


Figure 9. Gender Histograms for Reading

Mode	N	M	SD	T-test	Sig(2-tailed)
Male	227	16.99	5.31	12.941	.000
Female	212	24.08	6.09		

Table 3. Gender Means, Standard Deviations and T. Test Result for the Total English Test

	Mode	N	M	SD	T-test	Sig(2-tailed)
Grammar	Male	227	10.54	3.96	11.31	.000
	Female	212	15.17	4.61		
Reading	Male	227	6.44	2.53	9.42	.000
	Female	212	8.90	2.90		

Table 4. Gender Means, Standard Deviations and T. Test Result for the Grammar and Reading Test

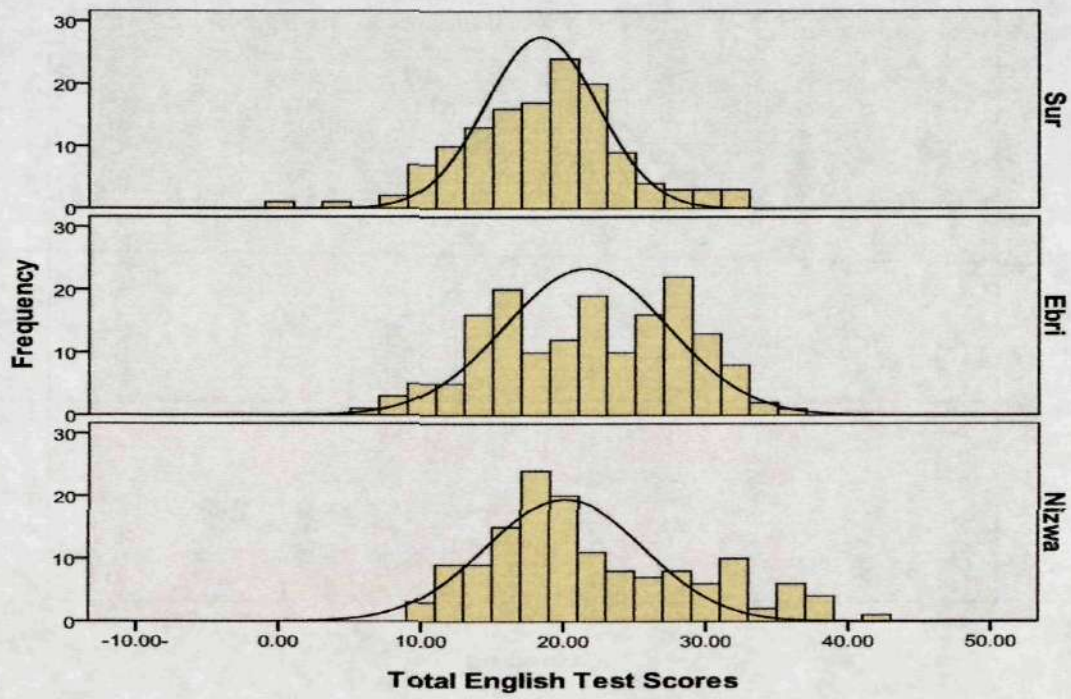


Figure 10. Colleges Histograms for theTotal English Test Scores

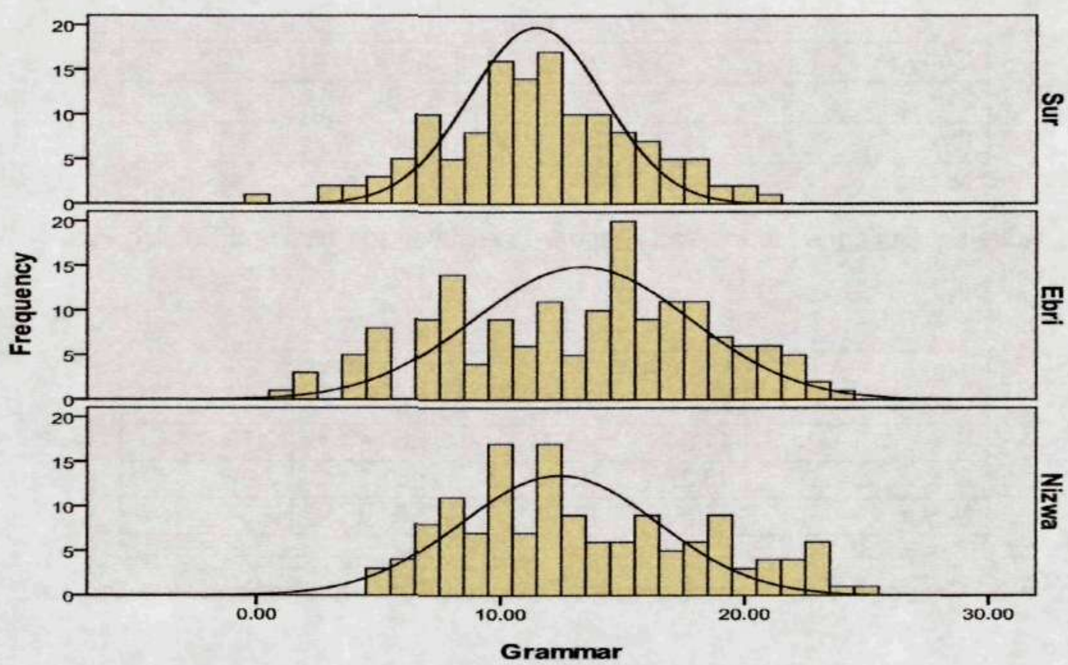


Figure 11. Colleges Histograms for Grammar

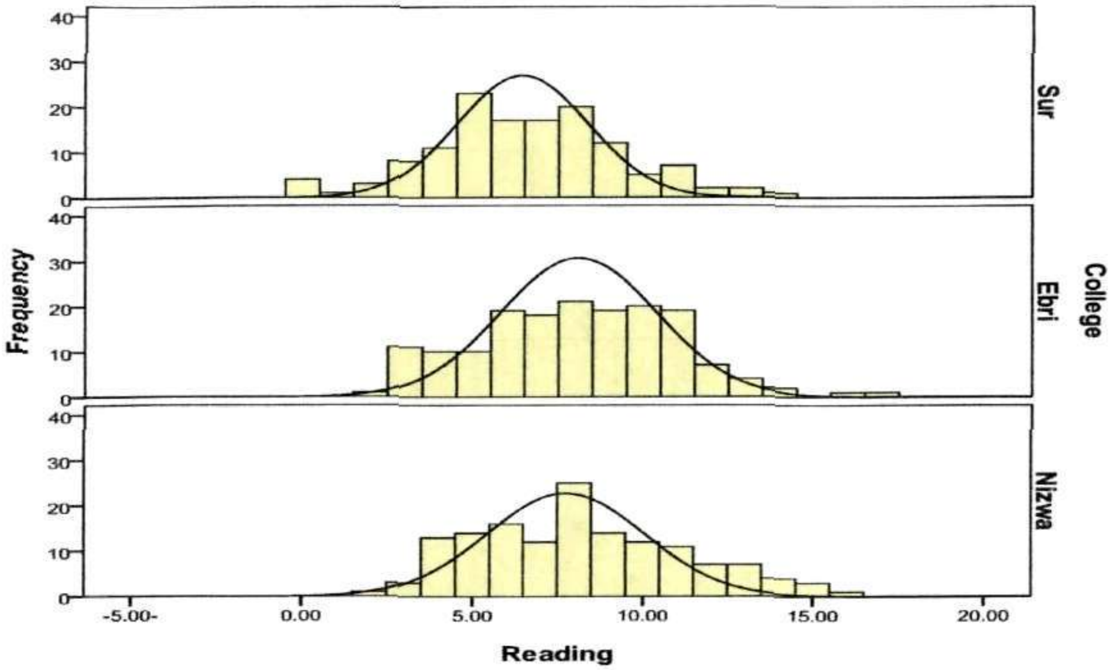


Figure 12. Colleges Histograms for Reading

College	Mean	SD
Sur	18.08	5.56
Ibri	21.28	6.66
Nizwa	21.58	7.23

Table 5. Colleges Means and Standard Deviation for the Total English Test

	Grammar		*Reading*	
College	M	SD	M	SD
Sur	11.54	3.94	6.54	2.76
Ibri	13.25	5.29	8.03	2.89
Nizwa	13.39	5.29	8.19	3.04

Table 6. Colleges Means and Standard Deviation For Grammar and Reading Test

Test of Homogeneity of Variances

	Levene Statistic	df1	df2	Sig.
TOTAL	7.277	2	436	.001
TOTGRAMM	9.752	2	436	.000
TOTREAD	.753	2	436	.472

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
TOTAL	Between Groups	1044.351	2	522.176	12.171	.000
	Within Groups	18706.195	436	42.904		
	Total	19750.547	438			
TOTGRAMM	Between Groups	293.788	2	146.894	6.344	.002
	Within Groups	10095.780	436	23.155		
	Total	10389.567	438			
TOTREAD	Between Groups	230.366	2	115.183	13.634	.000
	Within Groups	3683.319	436	8.448		
	Total	3913.686	438			

Table 7. ANOVA for the Total English Test, Grammar and Reading

College	Mean	Mean differences		
		Sur	Ibri	Nizwa
Sur	18.08	-	-3.20*	-3.50*
Ibri	21.28	3.20*	-	-0.29
Nizwa	21.58	3.50*	0.29	-

** The mean difference is significant at 0.05 level*

Table 8. Total English test Post Hock comparison for colleges

College	Mean	Mean differences		
		Sur	Ibri	Nizwa
Sur	11.54	-	-1.71*	-1.85*
Ibri	13.25	1.71*	-	-0.14
Nizwa	13.39	1.85*	0.14	-

**The mean difference is significant at 0.05 level*

Table 9. Grammar Post Hoc Comparison for Colleges

College	Mean	Mean differences		
		Sur	Ibri	Nizwa
Sur	6.54	-	-1.49*	-1.65*
Ibri	8.03	1.49*	-	-0.159
Nizwa	8.19	1.65*	0.159	-

** The mean difference is significant at 0.05 level*

Table 10. Reading Post Hoc Comparison for Colleges

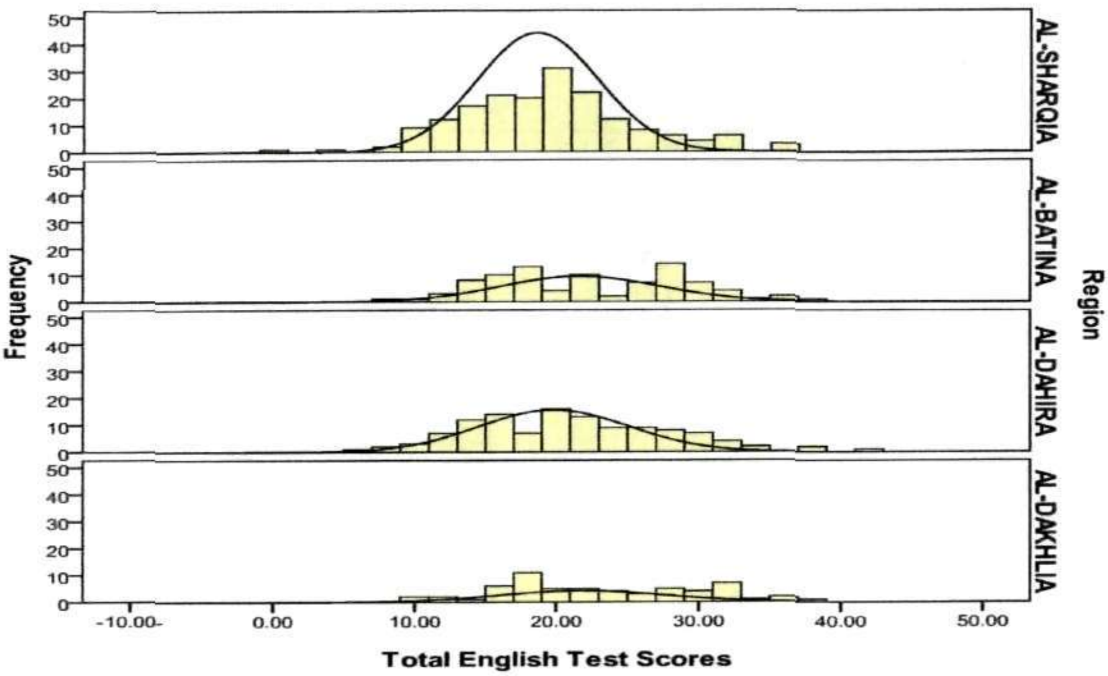


Figure 13. Regions Histograms for the Total English Test Scores

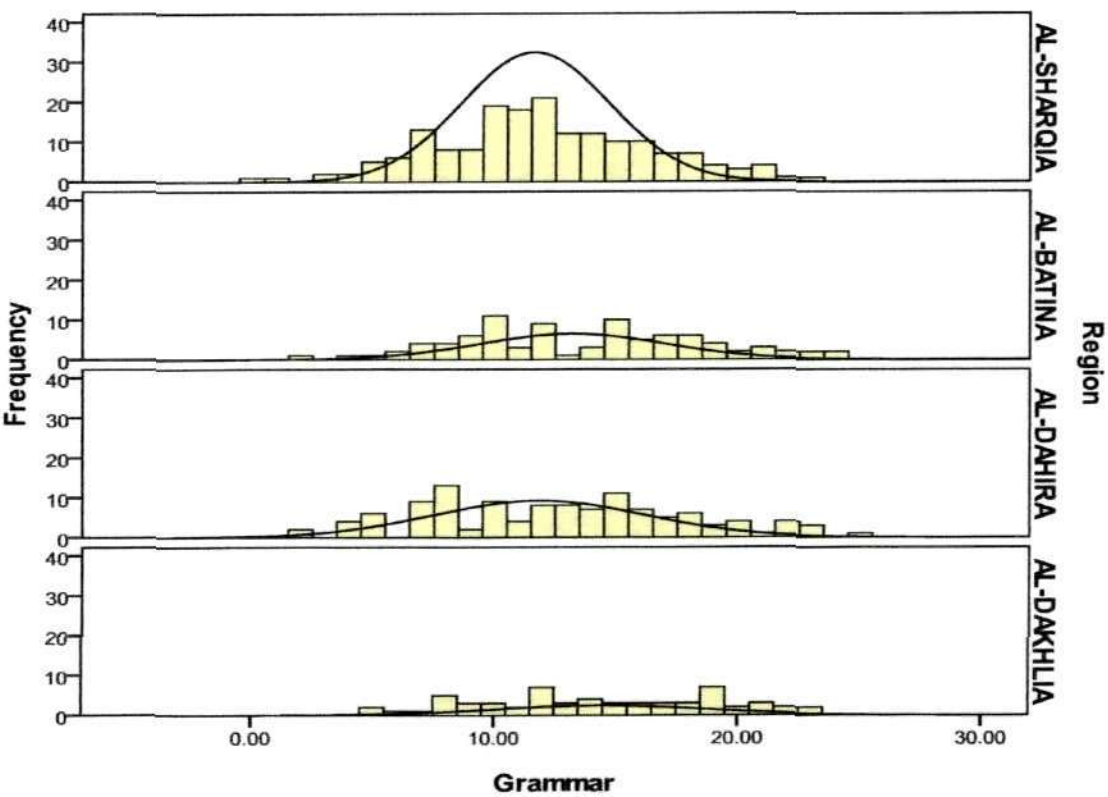


Figure 14. Regions Histograms for Grammar

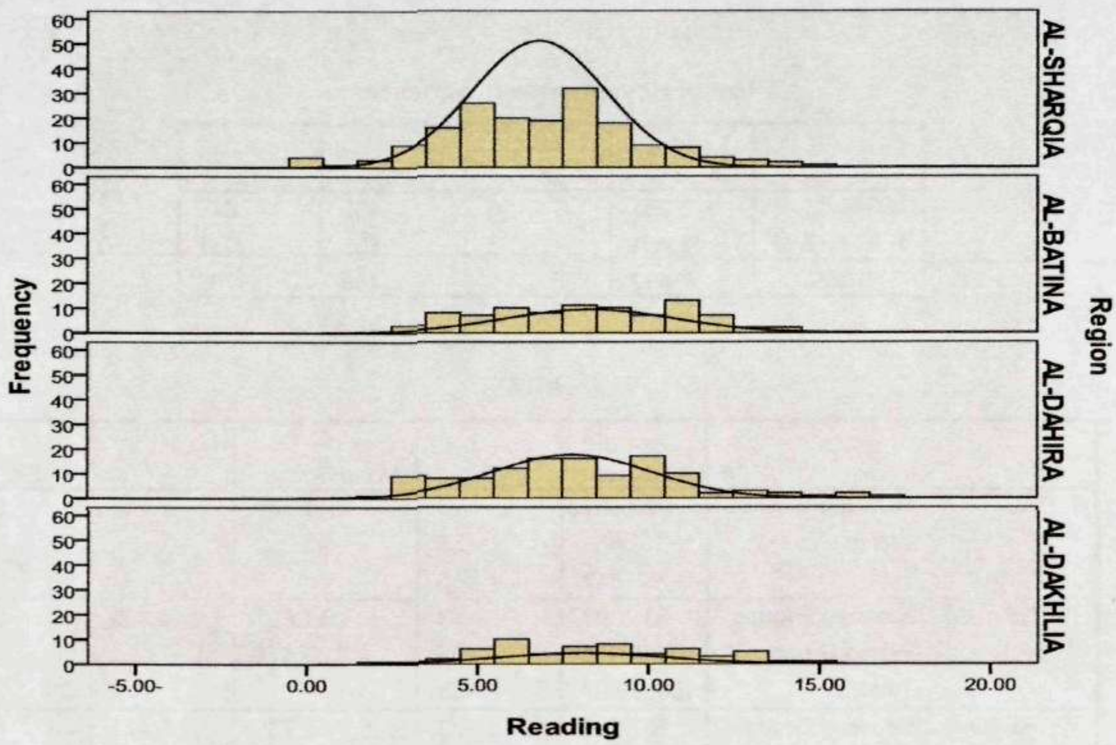


Figure 15. Regions Histograms for Reading

Regions	M	SD
Alsharqiya	18.82	6.11
AlBatinahh	21.82	6.69
AlDhahira	20.57	6.96
AlDakhlia	22.71	6.93

Table 11. Region Means and Standard Deviation for the Total English Test Scores

	Grammar		Reading	
Regions	M	SD	M	SD
Alsharqiya	11.92	4.35	6.89	2.83
AlBatinahh	13.63	4.98	8.19	2.83
AlDhahira	12.63	5.24	7.94	3.11
AlDakhlia	14.32	4.90	8.38	2.98

Table 12. Region Means and Standard Deviation for Grammar and Reading

Test of Homogeneity of Variances

	Levene Statistic	df1	df2	Sig.
TOTAL	2.281	3	435	.079
TOTGRAMM	3.454	3	435	.017
TOTREAD	.462	3	435	.709

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
TOTAL	Between Groups	933.861	3	311.287	7.196	.000
	Within Groups	18816.686	435	43.257		
	Total	19750.547	438			
TOTGRAMM	Between Groups	335.091	3	111.697	4.832	.003
	Within Groups	10054.476	435	23.114		
	Total	10389.567	438			
TOTREAD	Between Groups	167.206	3	55.735	6.471	.000
	Within Groups	3746.480	435	8.613		
	Total	3913.686	438			

Table 13. ANOVA for the Total English Test, Grammar and Reading

Region	Mean	Mean differences			
		SHR	DAK	DAH	BAT
SHR	18.82	-	-3.88*	-1.7	-3.00*
DAK	21.82	3.88*	-	2.13	0.88
DAH	20.57	1.74	-2.13	-	-1.25
BAT	22.71	3.00*	-0.88	1.25	-

**The mean difference is significant at 0.05 level*

Table 14. Total English Test Pos Hoc Comparison for Regions

Region	Mean	Mean differences			
		SHR	DAK	DAH	BAT
SHR	11.92	-	-2.39*	-0.70	-1.71*
DAK	13.63	2.39*	-	1.68	0.68
DAH	12.63	0.70	-1.68	-	-1.00
BAT	14.32	1.71*	-0.68	1.00	-

**The mean difference is significant at 0.05 level*

Table 15. Grammar Post Hoc Comparison for Regions

Region	Mean	Mean differences			
		SHR	DAK	DAH	BAT
SHR	6.89	-	-1.49*	-1.04	-1.29*
DAK	8.19	1.49*	-	0.44	0.19
DAH	7.94	1.04*	-0.44	-	-0.25
BAT	8.38	1.29*	-0.19	0.25	-

**The mean difference is significant at 0.05 level*

Table 16. Reading Post Hoc Comparison for Regions

	Mode	N	M	SD	T-test	Sig(2-tailed)
Male	Paper	132	16.3333	4.58785	2.116	.036
	Computer	95	17.9053	6.10384		
Female	Paper	111	25.5766	5.86135	3.885	.000
	Computer	101	22.4257	5.93691		

Table 17. Means, Standard Deviations and T. Test Result for the Gender/Mode Interaction on Total English Test Score

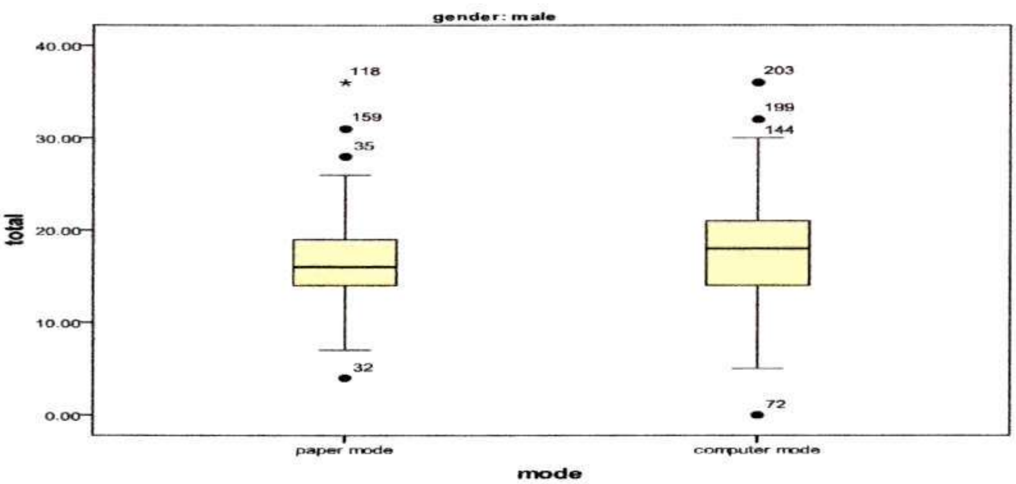


Figure 16. Mode Box Plots for Total English test Scores for Male

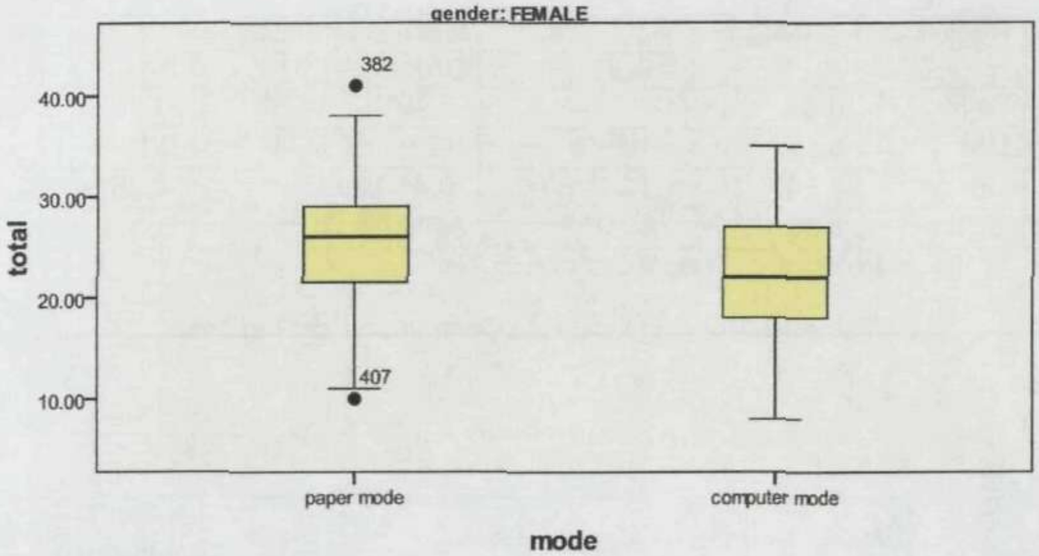


Figure 17. Mode Box Plots for the Total English Test Scores for Female

Group Statistics

Gender		Mode	N	Mean	Std. Deviation
Male	Grammar	paper mode	132	9.9318	3.12877
		computer mode	95	11.3789	4.78290
	Reading	paper mode	132	6.3939	2.60287
		computer mode	95	6.5263	2.45748
Female	Grammar	paper mode	111	16.1892	3.97838
		computer mode	101	14.0594	5.00764
	Reading	paper mode	111	9.3964	2.80026
		computer mode	101	8.3663	2.96217

Independent Samples Test

Gender			Levene's Test for Equality of Variances		t-test for Equality of Means		
			F	Sig.	T	df	Sig. (2-tailed)
Male	grammar	Equal variances assumed	23.732	.000	-2.754-	225	.006
		Equal variances not assumed			-2.579-	150.567	.011
	Reading	Equal variances assumed	.198	.657	-.387-	225	.699
		Equal variances not assumed			*-.391-*	*209.206*	.697
FEMALE	grammar	Equal variances assumed	2.763	.098	3.443	210	.001
		Equal variances not assumed			3.407	190.672	.001
	Reading	Equal variances assumed	.540	.463	2.602	210	.010
		Equal variances not assumed			2.595	205.328	.010

Table 18. Means, Standard Deviations and T. Test Result for the Gender/Mode Interaction on Grammar and Reading

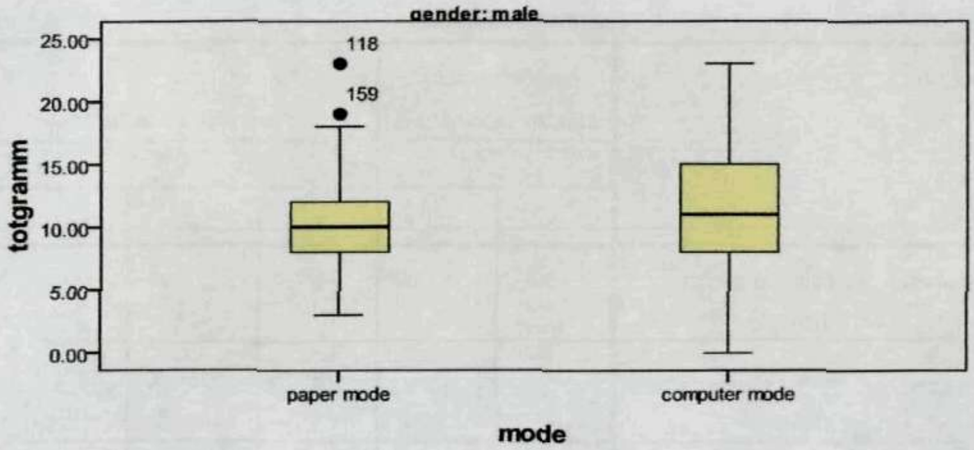


Figure 18. Grammar Mode Box Plots for Male

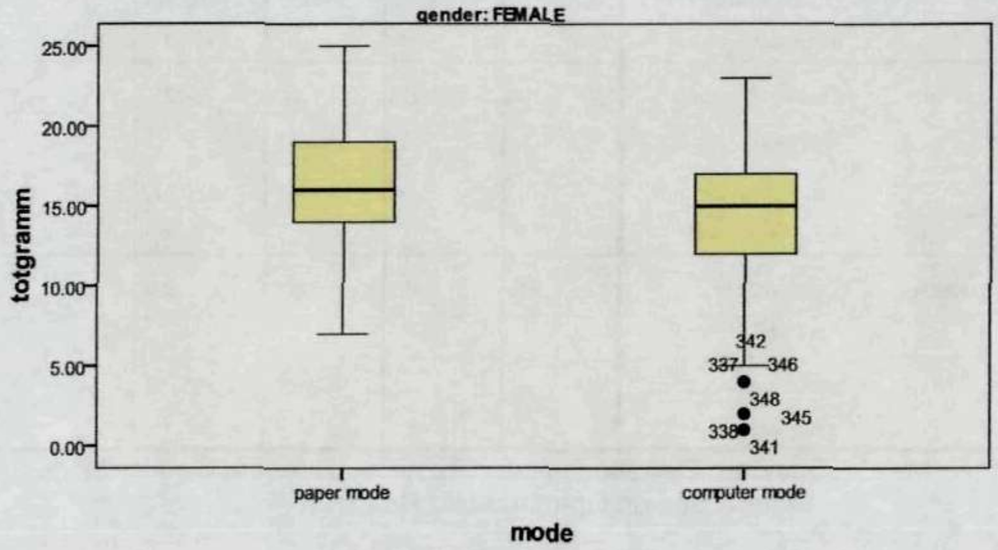


Figure 19. Grammar Mode Box Plots for Female

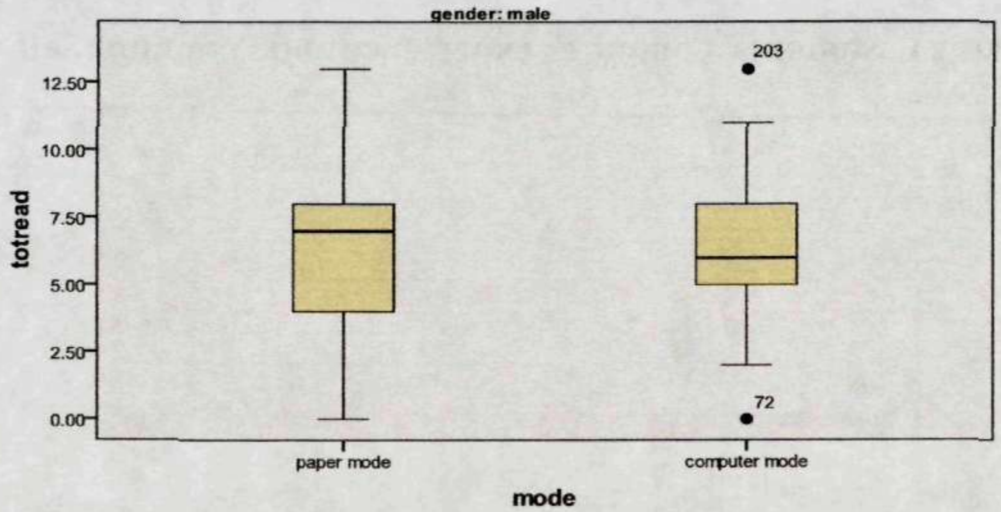


Figure 20. Reading Mode Box Plots for Male

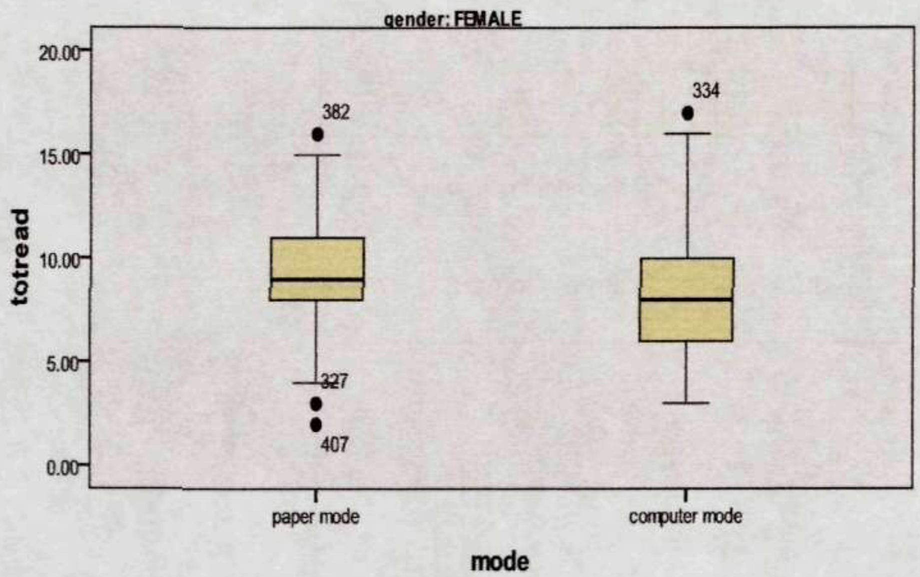


Figure 21. Reading Mode Box Plots for Female

Appendix C. Students' Computer Experience and Computer Self Efficacy

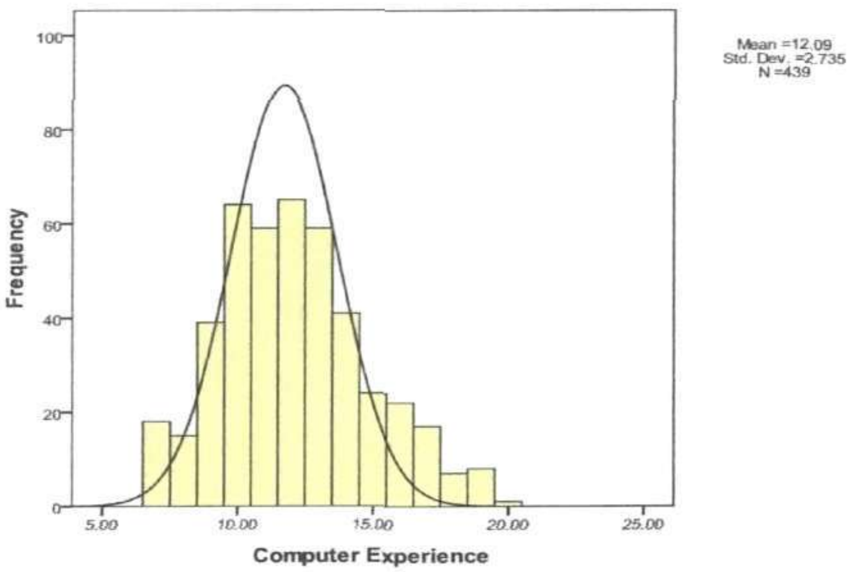


Figure 1. Histogram of Computer Experience

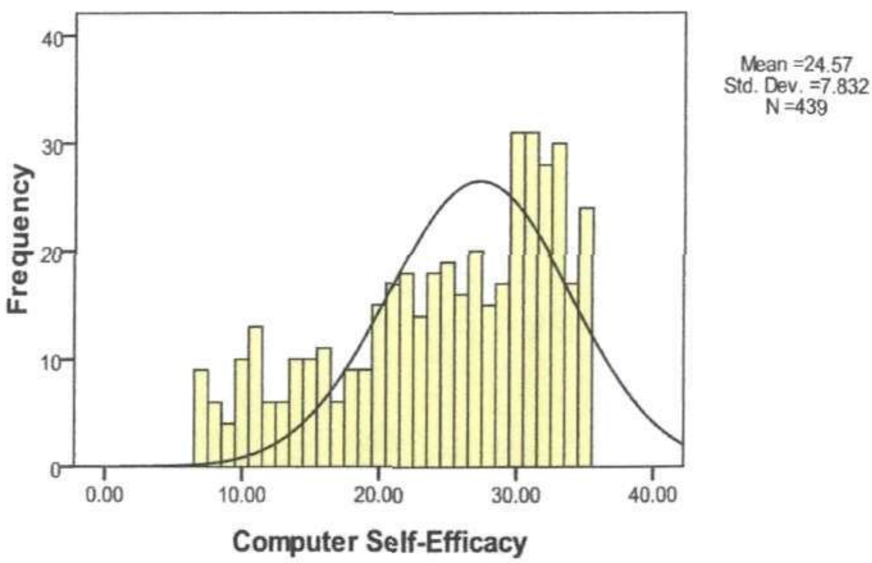


Figure 2. Histogram of Computer Self-Efficacy

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.759
Bartlett's Test of Sphericity	Approx. Chi-Square	495.096
	df	28
	Sig.	.000

Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.619	32.739	32.739	2.619	32.739	32.739
2	1.080	13.503	46.242			
3	1.018	12.727	58.969			
4	.885	11.063	70.032			
5	.734	9.179	79.211			
6	.627	7.836	87.047			
7	.547	6.838	93.884			
8	.489	6.116	100.000			

Extraction Method: Principal Component Analysis.

Component Matrix[a]

	Component
	1
I5	.729
I2	.702
I4	.641
I6	.637
I1	.601
I8	.467
I3	.430
I7	.123

Extraction Method: Principal Component Analysis.

a. 1 components extracted.

Table 1. Factor Analysis of Computer Experience

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.878
Bartlett's Test of Sphericity	Approx. Chi-Square	1528.161
	Df	21.000
	Sig.	.000

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.179	59.705	59.705	4.179	59.705	59.705
2	.841	12.019	71.724			
3	.525	7.493	79.217			
4	.496	7.086	86.303			
5	.376	5.370	91.673			
6	.325	4.649	96.323			
7	.257	3.677	100.000			

Extraction Method: Principal Component Analysis.

Component Matrix[a]

	Component
	1
it2	.829
it6	.780
it7	.777
it1	.771
it3	.769
it4	.768
it5	.710

Extraction Method:
Principal Component
Analysis.

a. 1 components
extracted.

Table 2. Factor Analysis of Computer Self-Efficacy

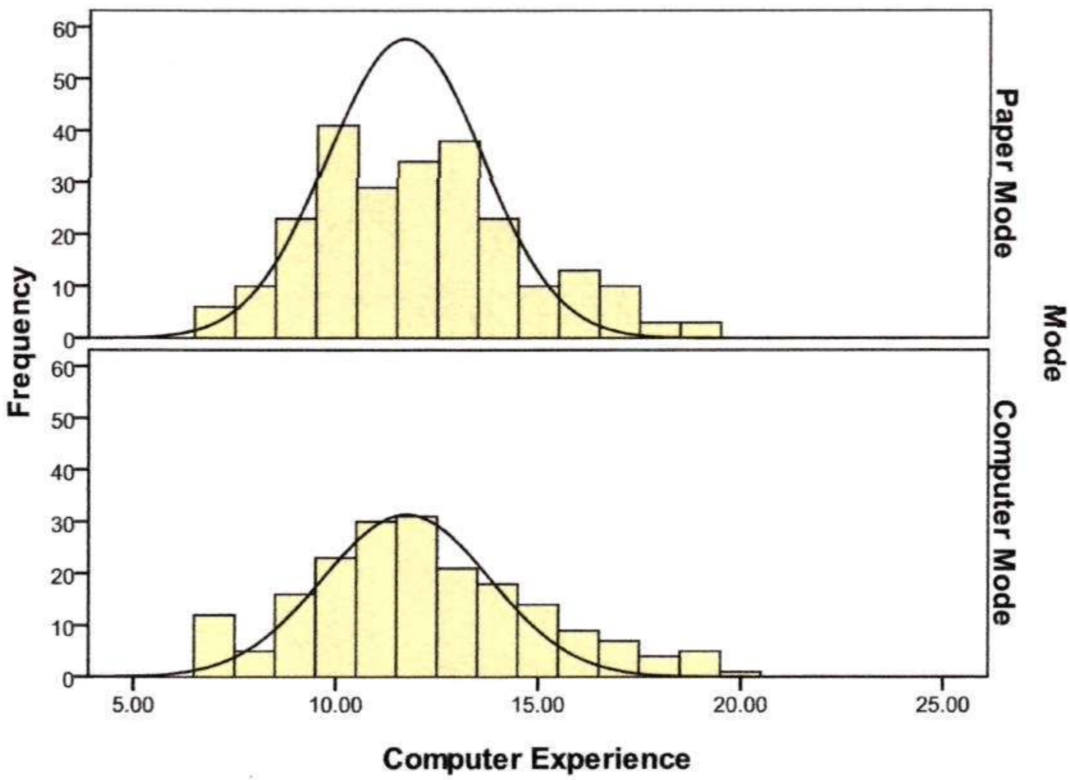


Figure 3. Mode Histograms for Computer Experience

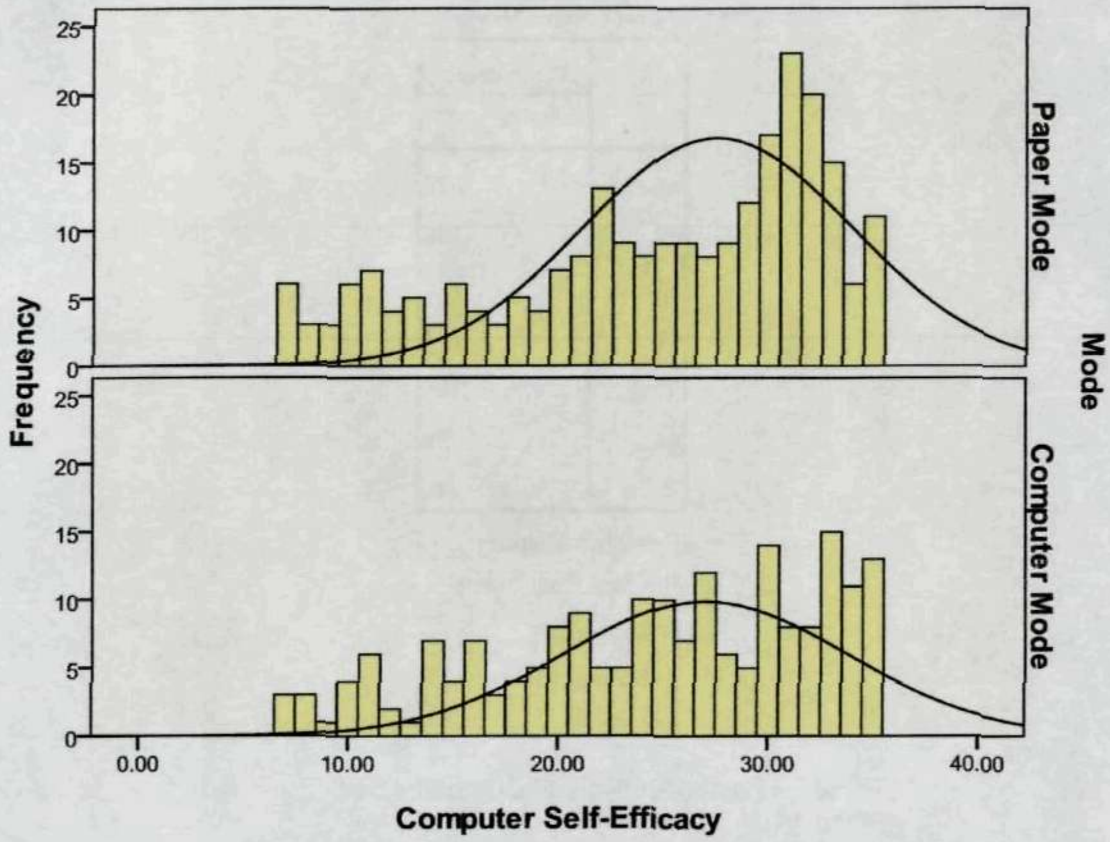


Figure 4. Mode Histograms for Computer Self -Efficacy

	Mode	N	M	SD	T-test	Sig(2-tailed)
Computer experience	Paper	243	12.02	2.61	.562	.574
	Computer	196	12.17	2.89		
Computer self-efficacy	Paper	243	24.75	7.89	.001	.999
	Computer	196	24.75	7.77		

Table 3. Means, Standard Deviations and T. Test Result for Computer Experience and Computer Self-Efficacy Regarding Mode of Assessment

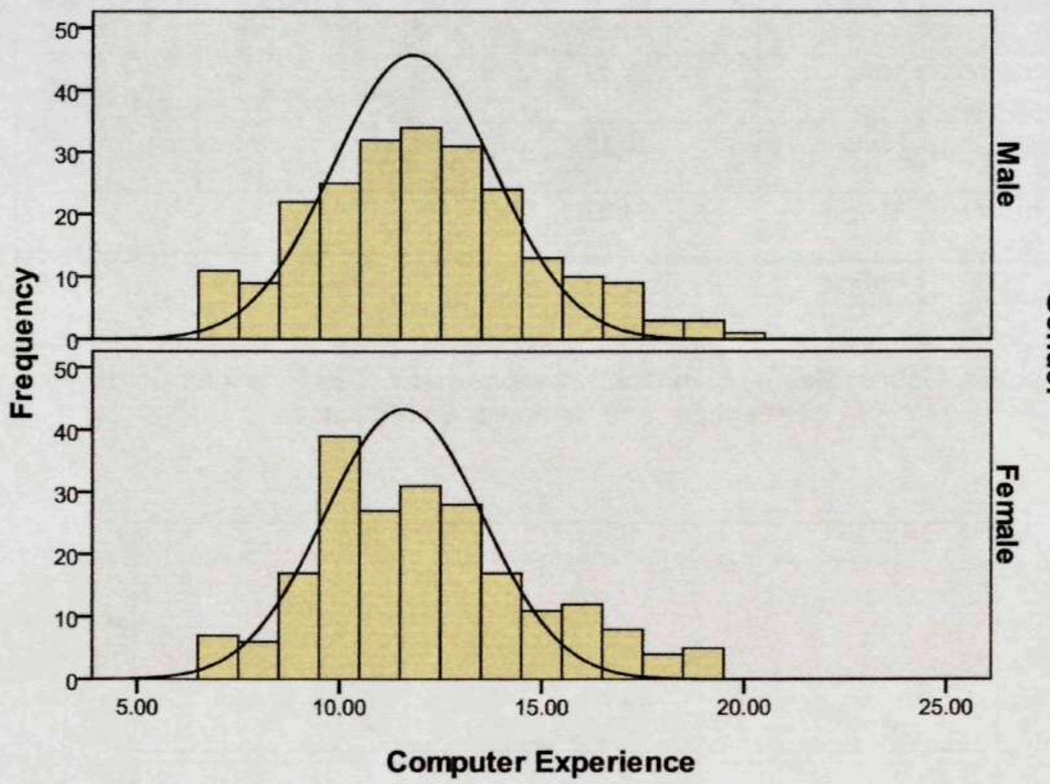


Figure 5. Gender Histograms for Computer Experience

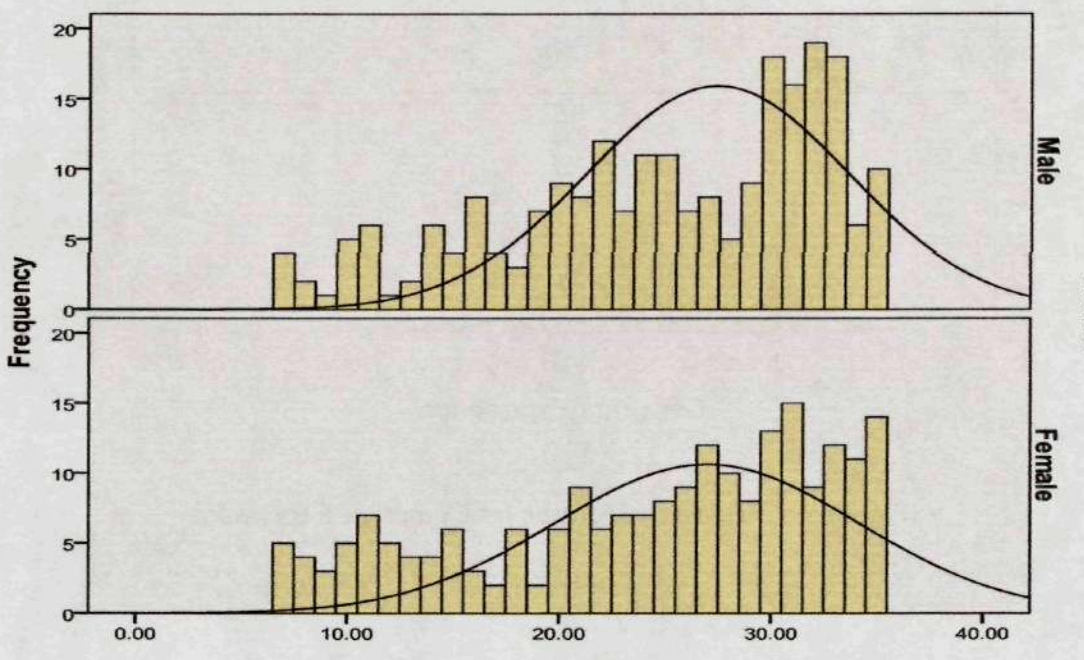


Figure 6. Gender Histograms for Computer Self-Efficacy

	Mode	N	M	SD	T-test	Sig(2-tailed)
Computer experience	Male		12.05	2.74	.302	.763
	Female		12.13	2.74		
Computer self-efficacy	Male		24.83	7.47	.722	.471
	Female		24.29	8.20		

Table 4. Gender Means, Standard Deviations and T. Test Result for Computer Experience and Computer Self-Efficacy

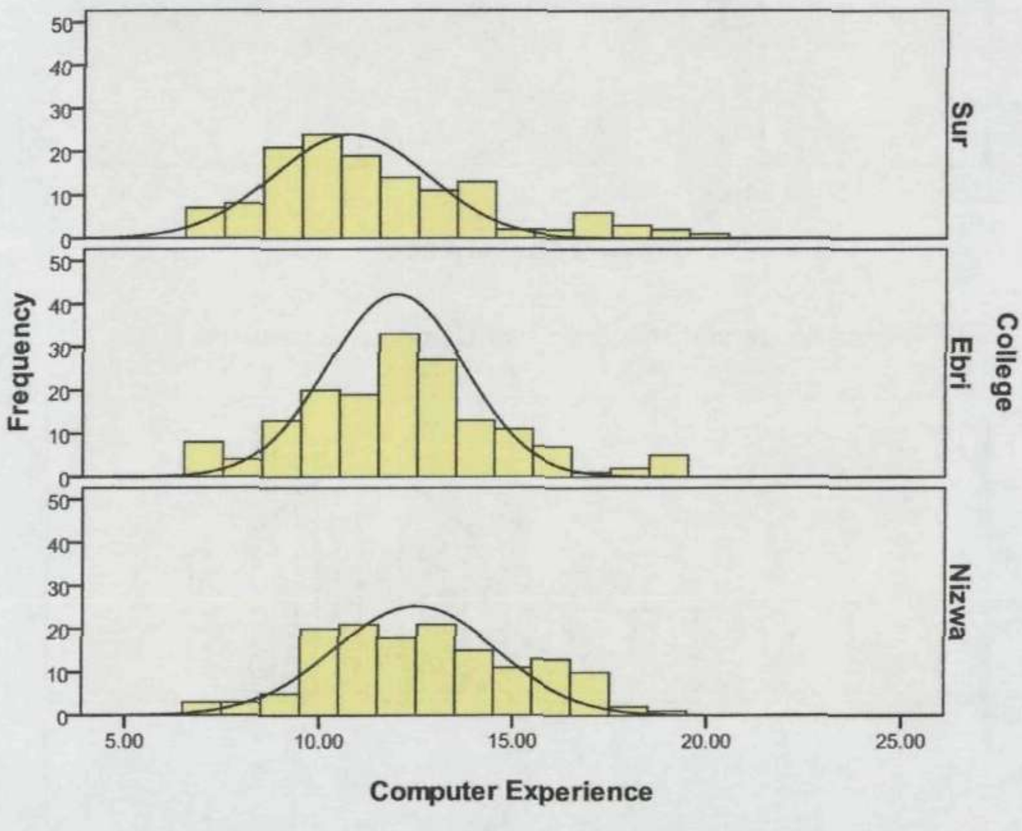


Figure 7. College Histograms for Computer Experience

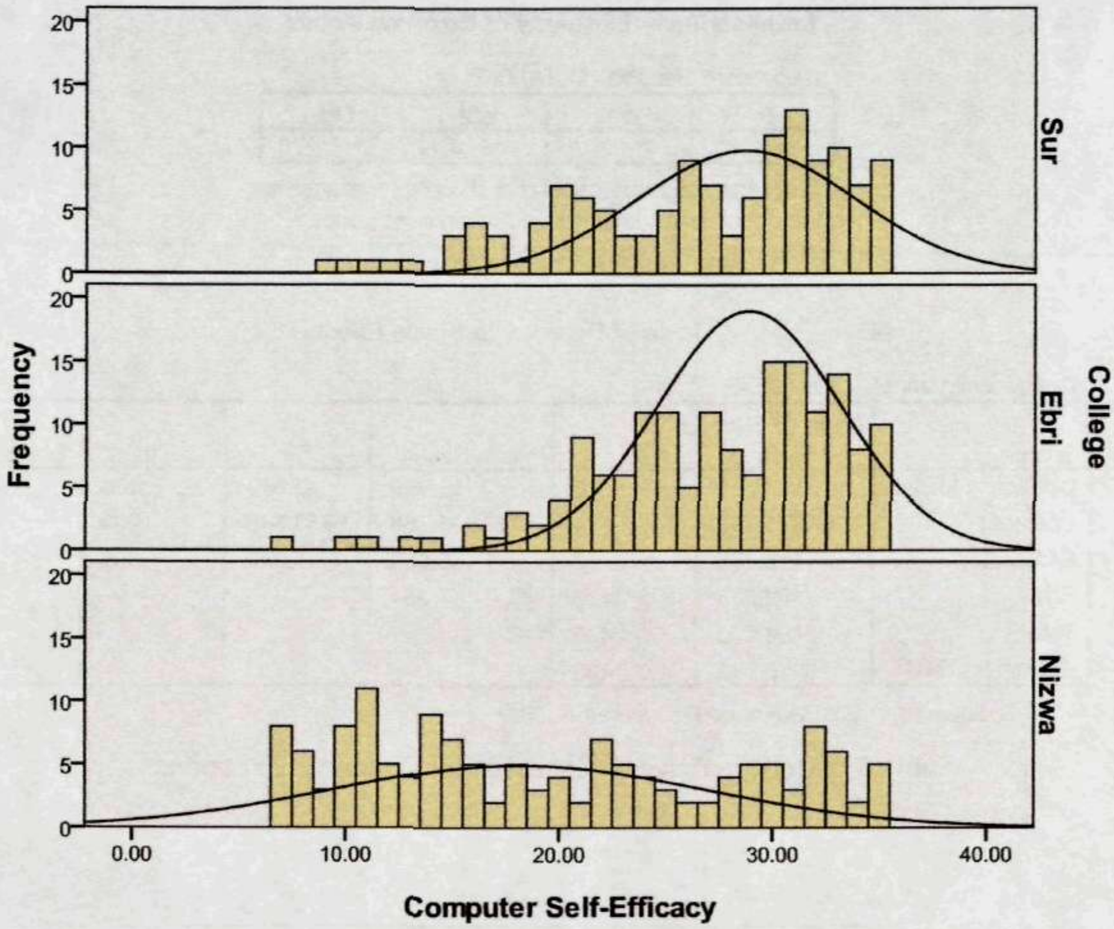


Figure 8. College Histograms for Computer Self-Efficacy

	Computer Experience		Computer Self-Efficacy	
College	M	SD	M	SD
Sur	11.43	2.88	26.58	6.42
Ibri	12.07	2.62	27.30	5.63
Nizwa	12.71	2.59	19.58	8.79

Table 5. Colleges' Means and Standard Deviation on Computer Experience and Computer Self-Efficacy

Levene's Test of Equality of Error Variances[a]

Dependent Variable: TOTALEXP

F	df1	df2	Sig.
1.283	2	436	.278

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept+COLLEGE

Tests of Between-Subjects Effects

Dependent Variable: TOTALEXP

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Eta Squared
Corrected Model	113.778[a]	2	56.889	7.842	.000	.035
Intercept	63523.768	1	63523.768	8756.545	.000	.953
COLLEGE	113.778	2	56.889	7.842	.000	.035
Error	3162.933	436	7.254			
Total	67408.000	439				
Corrected Total	3276.711	438				

a. R Squared = .035 (Adjusted R Squared = .030)

Table 6. College ANOVA test result on Computer Experience

College	Mean	Mean differences		
		Sur	Ibri	Nizwa
Sur	11.43	-	-0.645	-1.28*
Ibri	12.07	0.645	-	-0.639
Nizwa	12.71	1.28*	0.639	-

*The mean difference is significant at 0.000 level

Table 7. Colleges Post Hoc Tests Comparison on Computer Experience

Test of Homogeneity of Variances

Computer self efficacy

Levene Statistic	df1	df2	Sig.
30.156	2	436	.000

ANOVA

Computer Self –Efficacy

	Sum of Squares	Df	Mean Square	F	Sig.
Between Groups	5306.307	2	2653.153	53.656	.000
Within Groups	21559.183	436	49.448		
Total	26865.490	438			

Table 8. College ANOVA Test Result on Computer Self-Efficacy

College	Mean	Mean differences		
		Sur	Ibri	Nizwa
Sur	26.58	-	-0.714	6.99*
Ibri	27.30	-0.714	-	7.71*
Nizwa	19.58	-6.99*	-7.71*	-

*. The mean difference is significant at the 0.000 level.

Table 9. Colleges Post Hoc Tests Comparison on Computer Self-Efficacy

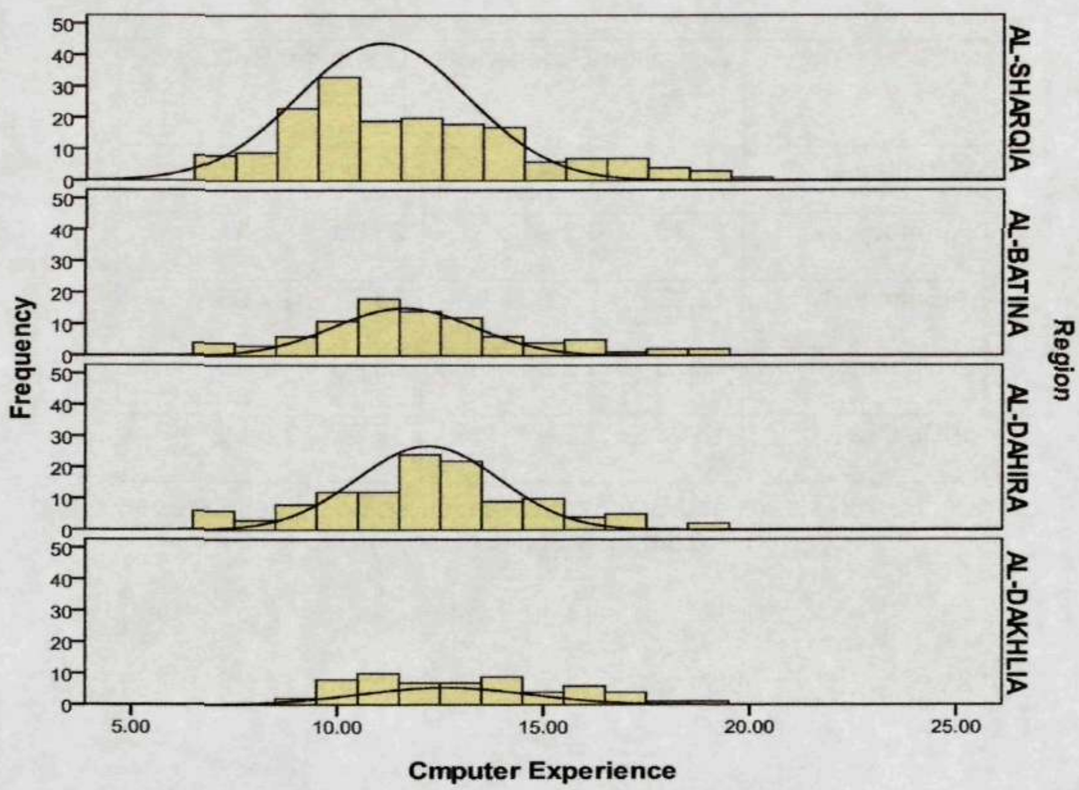


Figure 9. Region Histograms for Computer Experience

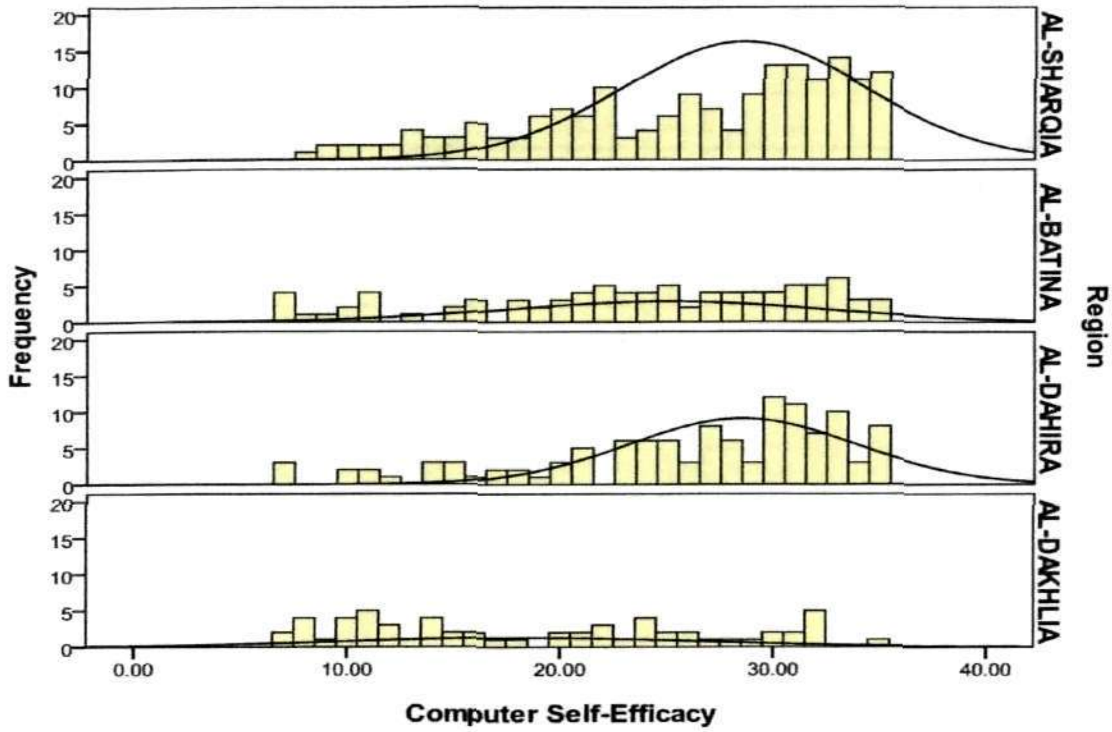


Figure 10. Region Histograms for Computer Self-Efficacy

	Computer Experience		Computer Self Efficacy	
Regions	M	SD	M	SD
AlSharqiya	11.75	2.91	25.80	7.17
AlBatinahH	11.97	2.65	23.71	7.95
AlDhahira	12.20	2.56	26.06	7.13
AlDakhlia	13.05	2.47	19.01	8.35

Table 10. Region Means and Standard Deviation on Computer Experience and Computer Self -Efficacy

Levene's Test of Equality of Error Variances[a]

Dependent Variable: TOTALEXP

F	df1	df2	Sig.
1.782	3	435	.150

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept+REGION

Tests of Between-Subjects Effects

Dependent Variable: TOTALEXP

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Eta Squared
Corrected Model	77.550[a]	3	25.850	3.515	.015	.024
Intercept	56308.166	1	56308.166	7656.399	.000	.946
REGION	77.550	3	25.850	3.515	.015	.024
Error	3199.161	435	7.354			
Total	67408.000	439				
Corrected Total	3276.711	438				

a. R Squared = .024 (Adjusted R Squared = .017)

Table 11. Region ANOVA Test Result on Computer Experience

REGION

Dependent Variable: TOTALEXP

REGION	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
SHARQIA	11.749	.205	11.346	12.151
BITINA	11.966	.289	11.398	12.534
DAHIRA	12.197	.251	11.704	12.689
DAKHLIA	13.051	.353	12.357	13.745

Table 12. Region Estimated Marginal Means on Computer Experience

Region	Mean	Mean differences			
		SHR	DAK	DAH	BAT
SHR	11.75	-	-1.30*	-0.448	-0.217
DAK	11.97	1.30*	-	0.854	1.08
DAH	12.20	0.448	-0.854	-	0.450
BAT	13.05	0.217	-1.08	-0.045	-

*The mean difference is significant at 0.01 level

Table 13. Region Post Hoc Tests Comparison on Computer Experience

Test of Homogeneity of Variances

Computer self efficacy

Levene Statistic	df1	df2	Sig.
2.478	3	435	.061

ANOVA

Computer Self –Efficacy

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2443.922	3	814.641	14.510	.000
Within Groups	24421.568	435	56.142		
Total	26865.490	438			

Table 14. Region ANOVA Test Result on Computer Self-Efficacy

Test of Homogeneity of Variances

Computer Experience

Region	Levene Statistic	df1	df2	Sig.
SHARQYIA	.049	2	172	.952
BATINAH	2.974	1	86	.088
DHAHIRA	.500	1	115	.481
DAKHLIA	1.005	1	57	.320

ANOVA

Computer Experience

Region		Sum of Squares	df	Mean Square	F	Sig.
SHARQYIA	Between Groups	91.966	2	45.983	5.727	.004
	Within Groups	1380.971	172	8.029		
	Total	1472.937	174			
BATINAH	Between Groups	66.825	1	66.825	10.563	.002
	Within Groups	544.073	86	6.326		
	Total	610.898	87			
DHAHIRA	Between Groups	3.389	1	3.389	.515	.475
	Within Groups	757.090	115	6.583		
	Total	760.479	116			
DAKHLIA	Between Groups	9.567	1	9.567	1.579	.214
	Within Groups	345.280	57	6.058		
	Total	354.847	58			

Table 15. ANOVA Result for the College /Region Interaction.

Region	Mean	Mean differences			
		SHR	DAK	DAH	BAT
SHR	25.80	-	6.85*	-0.19	2.15
DAK	23.71	6.85*	-	7.05*	-4.69*
DAH	26.06	0.19	-7.05*	-	2.35
BAT	19.01	-2.15	4.69*	-2.35	-

*. The mean difference is significant at the 0.05 level.

Table. Region Post Hoc Tests Comparison on Computer Self-Efficacy

Test of Homogeneity of Variances

Computer Self-Efficacy

college	Levene Statistic	df1	df2	Sig.
Sur	.528	1	131	.469
Ibri	.221	1	161	.639
Nizwa	1.394	1	141	.240

ANOVA

SELFEFFI

COLLEGE		Sum of Squares	df	Mean Square	F	Sig.
sur	Between Groups	109.689	1	109.689	2.693	.103
	Within Groups	5336.566	131	40.737		
	Total	5446.256	132			
ebri	Between Groups	15.115	1	15.115	.475	.492
	Within Groups	5125.155	161	31.833		
	Total	5140.270	162			
nizwa	Between Groups	1038.233	1	1038.233	14.7	.000
	Within Groups	9934.425	141	70.457		
	Total	10972.657	142			

Table 16. ANOVA Result of Gender/College Interaction in Computer Self -Efficacy.

college	Mean	Mean differences		
		Sur	Ibri	Nizwa
Sur	*12.98*	-	-3.200*	-1.125
Ibri	*16.67*	3.200	-	2.075
Nizwa	*14.31*	-1.125	-2.075	-

*. The mean difference is significant at the 0.05 level

Table 17. Computer Experience Post Hoc Comparison for AlSharqiya Region Students Studying at Three Colleges

Appendix D. Student Attitudes Towards Computerised Assessment (ATCAS)

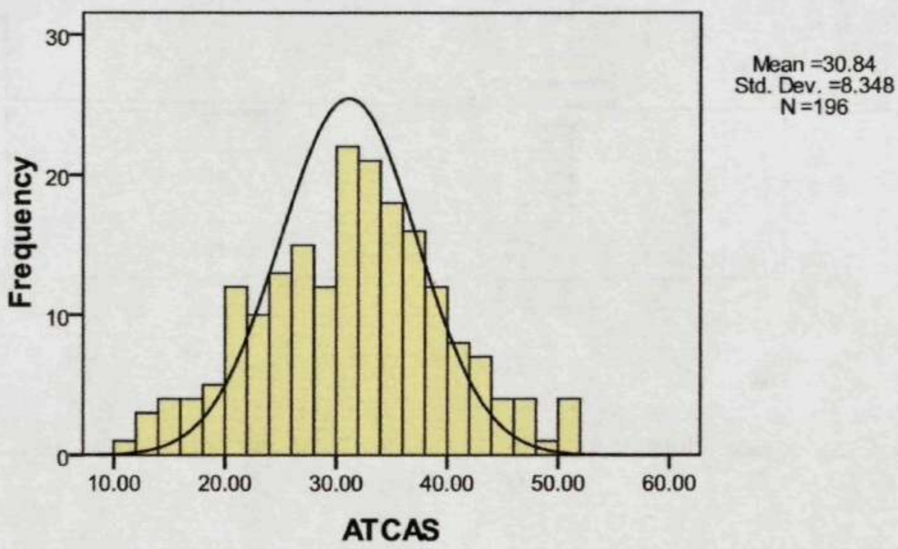


Figure 1. Histograms of Attitudes Towards Computerised Assessment Scale (ATCAS)

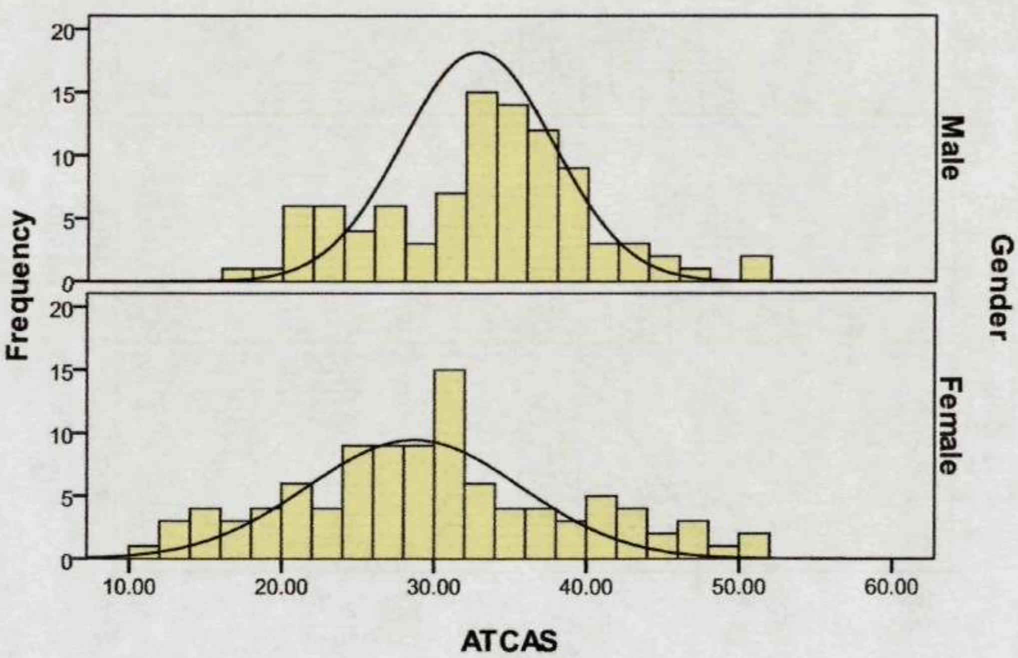


Figure 2. Gender Histograms of Attitudes Towards Computerised Assessment Scale(ATCAS)

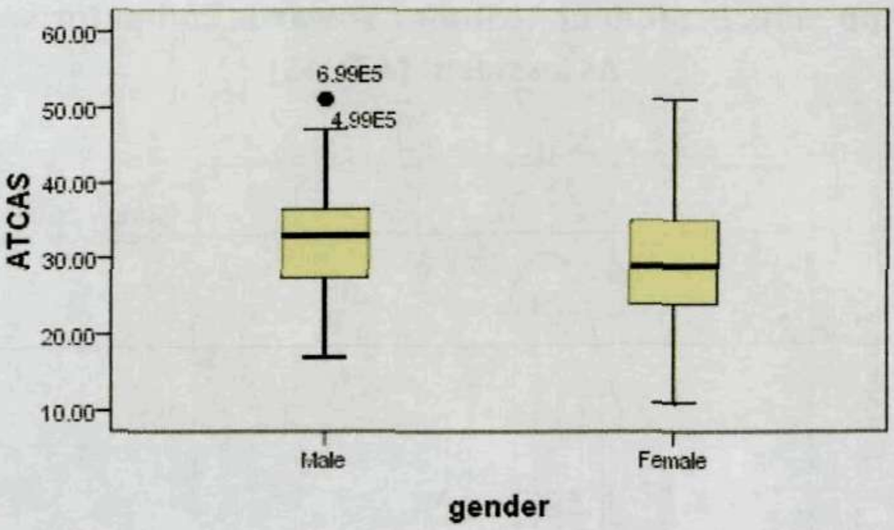


Figure 3. Gender Box Plots of Attitudes Towards Computerised Assessment Scale (ATCAS)

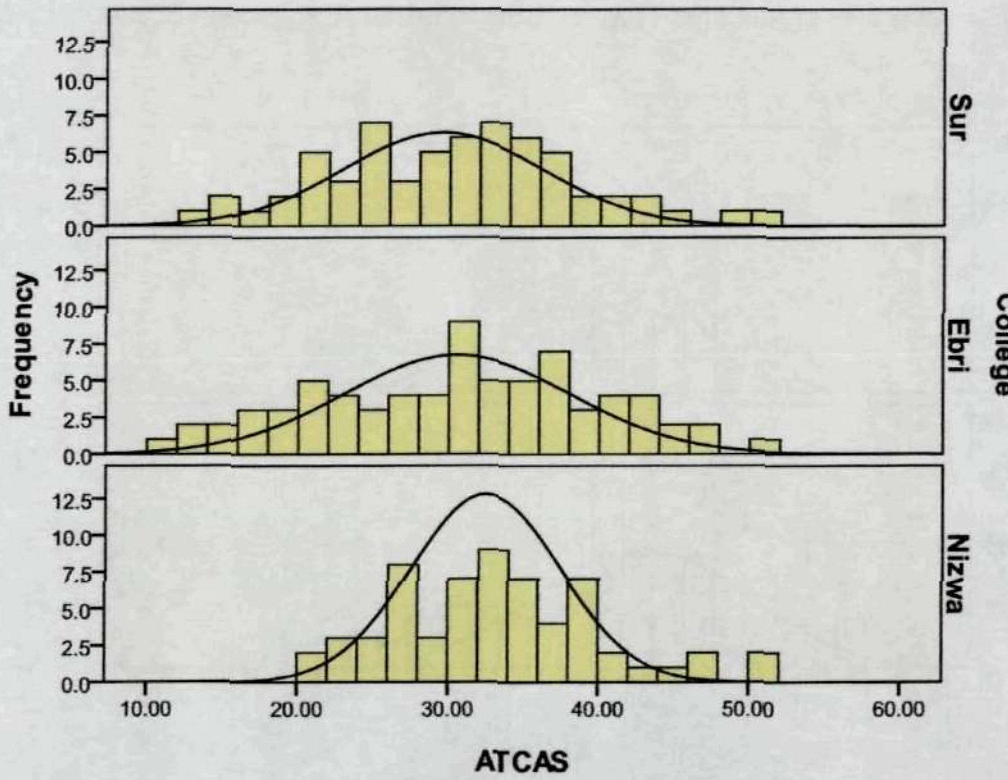


Figure 4. College Histogram of Attitudes Towards Computerised Assessment Scale (ATCAS)

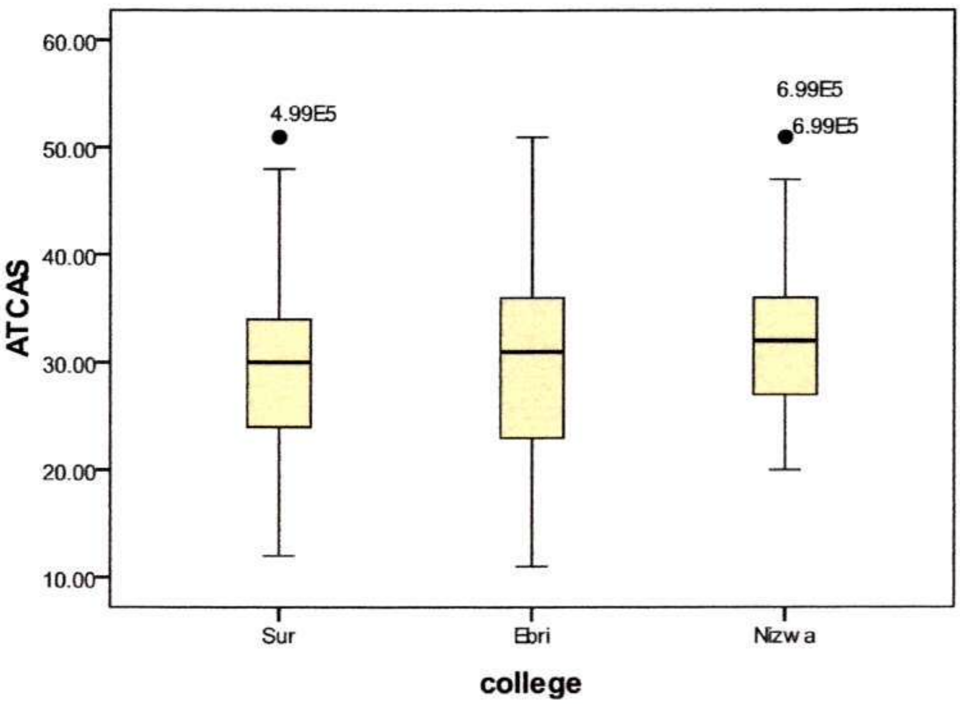


Figure 5. College Box Plots of Attitudes Towards Computerised Assessment Scale (ATCAS)

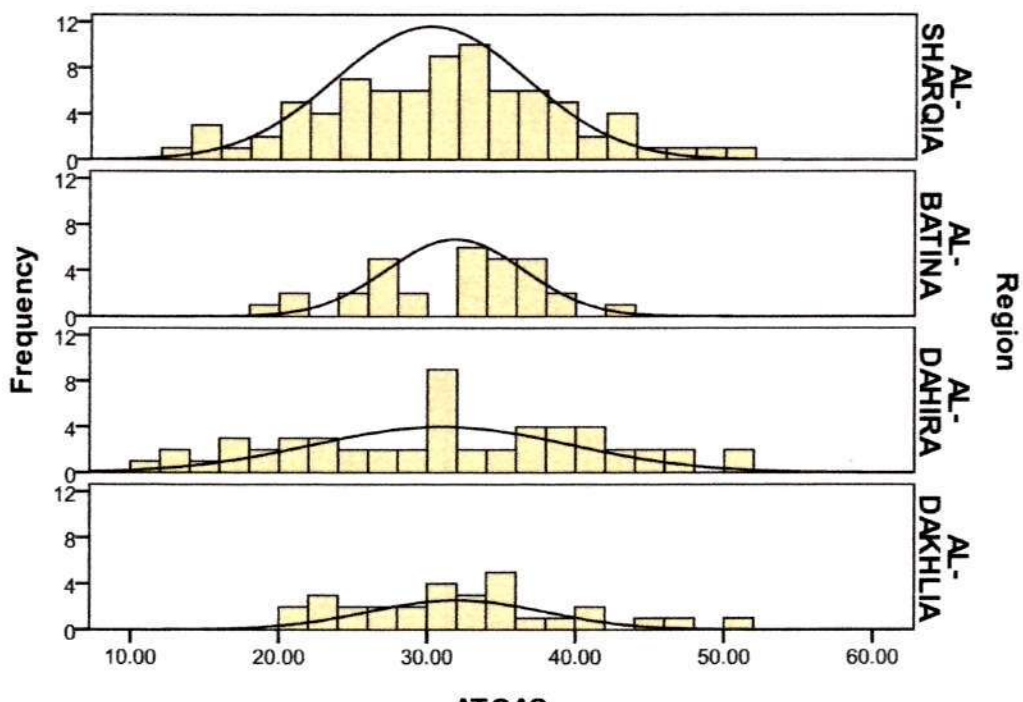


Figure 6. Region Histogram of Attitudes Towards Computerised Assessment Scale (ATCAS)

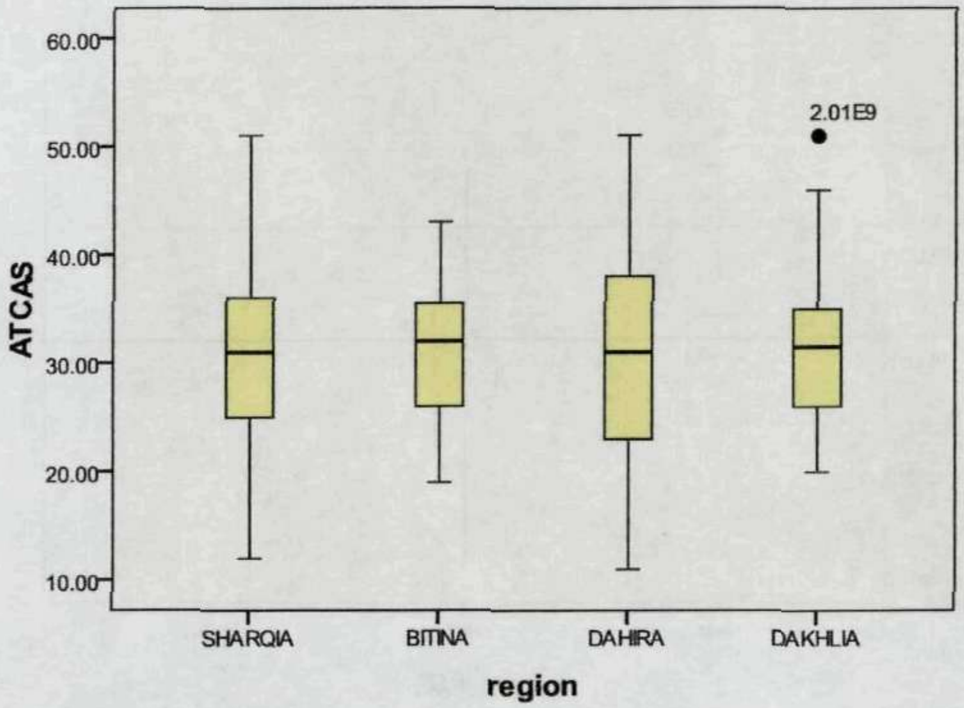


Figure 7. Region Box Plots of Attitudes Towards Computerised Assessment Scale (ATCAS)

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.751
Bartlett's Test of Sphericity	Approx. Chi-Square	507.986
	df	78
	Sig.	.000

Communalities

	Initial	Extraction
Q1	1.000	.418
Q2	1.000	.278
Q3	1.000	.346
Q4	1.000	.234
Q5	1.000	6.231E-02
Q6	1.000	.478
Q7	1.000	.163
Q8	1.000	2.660E-02
Q9	1.000	.238
Q10	1.000	.188
Q11	1.000	.354
Q12	1.000	.422
Q13	1.000	.319

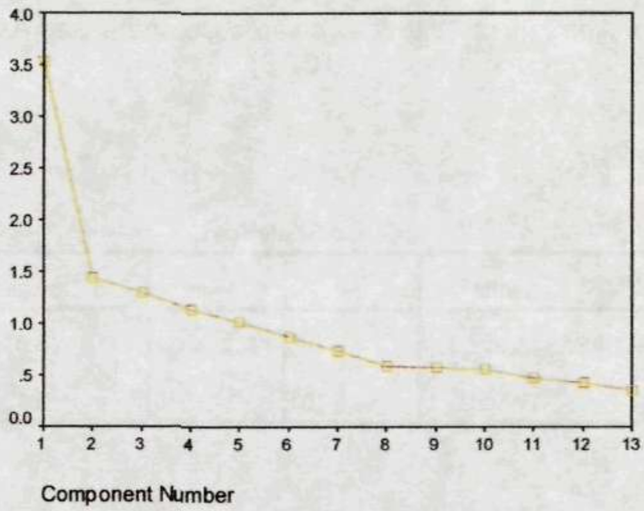
Extraction Method: Principal Component Analysis.

Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.526	27.121	27.121	3.526	27.121	27.121
2	1.447	11.133	38.254			
3	1.301	10.006	48.260			
4	1.131	8.701	56.962			
5	1.009	7.761	64.723			
6	.873	6.715	71.438			
7	.741	5.703	77.141			
8	.590	4.539	81.680			
9	.568	4.367	86.047			
10	.561	4.312	90.359			
11	.467	3.590	93.949			
12	.428	3.295	97.244			
13	.358	2.756	100.000			

Extraction Method: Principal Component Analysis.

Scree Plot



Component Matrix[a]

	Component
	1
Q6	.691
Q12	.649
Q1	.646
Q11	.595
Q3	.588
Q13	.565
Q2	.528
Q9	.488
Q4	.483
Q10	.433
Q7	.404
Q5	.250
Q8	.163

Extraction Method: Principal Component Analysis.

a. 1 components extracted.

Table 1. Factor Analysis of Attitudes Towards Computerised Assessment Scale (ATCAS)

Group Statistics

	gender	N	Mean	Std. Deviation
ATCAS	Male	95	32.4632	7.03912
	Female	101	29.3069	9.18993

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means			
		F	Sig.	T	Df	Sig. (2-tailed)	Mean Difference
ATCAS	Equal variances assumed	5.673	.018	2.687	194	.008	3.15623
	Equal variances not assumed			2.709	186.475	.007	3.15623

Table 2. Independent T-Test Result for Gender Differences on ATCAS

Test of Homogeneity of Variances

ATCAS

Levene Statistic	df1	df2	Sig.
3.079	2	193	.048

ANOVA

ATCAS

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	331.566	2	165.783	2.413	.092
Within Groups	13259.210	193	68.701		
Total	13590.776	195			

Table 3. ANOVA Test Result for College Difference on ATCAS

Test of Homogeneity of Variances

ATCAS

Levene Statistic	df1	df2	Sig.
3.409	3	192	.019

ANOVA

ATCAS

	Sum of Squares	Df	Mean Square	F	Sig.
Between Groups	52.482	3	17.494	.248	.863
Within Groups	13538.294	192	70.512		
Total	13590.776	195			

Table 4. ANOVA Test Result for Region Difference on ATCAS

Tests of Between-Subjects Effects

Dependent Variable: Grammar

Source	Type III Sum of Squares	Df	Mean Square	F	Sig.
Corrected Model	949.706[a]	20	47.485	2.047	.007
Intercept	384.716	1	384.716	16.583	.000
Tatcas	26.605	1	26.605	1.147	.286
Eperience	6.567	1	6.567	.283	.595
self-efficacy	6.632	1	6.632	.286	.594
Gender	45.129	1	45.129	1.945	.165
Region	29.499	3	9.833	.424	.736
College	29.437	2	14.718	.634	.531
gender * region	103.954	3	34.651	1.494	.218
gender * college	48.943	2	24.472	1.055	.350
region * college	148.632	3	49.544	2.136	.097
gender * region * college	156.957	3	52.319	2.255	.084
Error	4060.024	175	23.200		
Total	36923.000	196			
Corrected Total	5009.730	195			

a. R Squared = .190 (Adjusted R Squared = .097)

Table 5. Multi Factor ANCOVA Between ATCAS on Grammar

Tests of Between-Subjects Effects

Dependent Variable: Reading

Source	Type III Sum of Squares	Df	Mean Square	F	Sig.
Corrected Model	335.595[a]	20	16.780	2.303	.002
Intercept	178.129	1	178.129	24.444	.000
ATCAS	1.678	1	1.678	.230	.632
Experience	3.565	1	3.565	.489	.485
Self efficacy	.018	1	.018	.002	.961
Gender	90.644	1	90.644	12.439	.001
Region	11.000	3	3.667	.503	.681
College	8.240	2	4.120	.565	.569
gender * region	1.547	3	.516	.071	.976
gender * college	.813	2	.406	.056	.946
region * college	78.654	3	26.218	3.598	.015
gender * region * college	5.273	3	1.758	.241	.868
Error	1275.278	175	7.287		
Total	12561.000	196			
Corrected Total	1610.872	195			

a. R Squared = .208 (Adjusted R Squared = .118)

Table 6. Multi Factor ANCOVA Between ATCAS on Reading

Test of Homogeneity of Variances

Total

Region	Levene Statistic	df1	df2	Sig.
SHARQIA	3.155	2	172	.045
BITINA	.236	1	86	.628
DAHIRA	.846	1	115	.360
DAKHLIA	.199	1	57	.657

ANOVA

Total

Region		Sum of Squares	Df	Mean Square	F	Sig.
SHARQIA	Between Groups	489.800	2	244.900	7.000	.001
	Within Groups	6017.709	172	34.987		
	Total	6507.509	174			
BITINA	Between Groups	326.012	1	326.012	7.904	.006
	Within Groups	3547.079	86	41.245		
	Total	3873.091	87			
DAHIRA	Between Groups	74.992	1	74.992	1.553	.215
	Within Groups	5553.640	115	48.293		
	Total	5628.632	116			
DAKHLIA	Between Groups	40.582	1	40.582	.843	.363
	Within Groups	*2745.520*	*57*	*48.167*		
	Total	2786.102	58			

Table 7. ANOVA Test Result for Region/College Interaction on Total English Test Score

Multiple Comparisons

Total

Dunnett C

Region	(I) college	(J) college	Mean Difference (I-J)	Std. Error
SHARQIA	*Sur*	*Ibri*	*.68271*	*1.77428*
		Nizwa	-4.26104-*	1.38536
	Ibri	Sur	-.68271-	1.77428
		Nizwa	-4.94375-	2.14529
	Nizwa	Sur	4.26104*	1.38536
		Ibri	4.94375	2.14529

*. The mean difference is significant at the 0.05 level.

Table 8. Post Hoc Comparisons for Region/College Interaction on Total English Test Score

Test of Homogeneity of Variances

Reading

Region	Levene Statistic	df1	df2	Sig.
SHARQIA	.943	2	172	.391
BITINA	.340	1	86	.561
DAHIRA	.037	1	115	.847
DAKHLIA	.769	1	57	.384

ANOVA

Reading

Region		Sum of Squares	Df	Mean Square	F	Sig.
SHARQIA	Between Groups	81.351	2	40.676	5.305	.006
	Within Groups	1318.798	172	7.667		
	Total	1400.149	174			
BITINA	Between Groups	11.461	1	11.461	1.432	.235
	Within Groups	688.255	86	8.003		
	Total	699.716	87			
DAHIRA	Between Groups	13.055	1	13.055	1.346	.248
	Within Groups	1115.526	115	9.700		
	Total	1128.581	116			
DAKHLIA	Between Groups	7.358	1	7.358	.821	.369
	Within Groups	510.676	57	8.959		
	Total	518.034	58			

Table 9. ANOVA Test Result for Region/College Interaction on Reading

Tukey HSD

College	Mean	Mean differences		
		Sur	Ibri	Nizwa
Sur	6.54	-	-.55865	-1.77115-*
Ibri	7.100	.55865	-	-1.21250-
Nizwa	8.31	1.77115*	1.21250	-

*. The mean difference is significant at the 0.05 level

Table 10. Post Hoc Comparison for Region/College Interaction on Reading

(1) Strongly Disagree (2) (Disagree 3) Moderate (4) Agree (5) Strongly Agree

	Statements	**1**	**2**	**3**	**4**	**5**
1	More nervous on computer than paper	18.4 (36)	28.1 (55)	26.0 (51)	11.7 (23)	15.8 (31)
2	Test instructions on computer were difficult to understand	19.9 (39)	23.5 (46)	25.5 (50)	17.3 (34)	13.8 (27)
3	Helpful if given more practice time on the computer before starting the test	11.2 (22)	6.6 (13)	15.3 (30)	25.5 (50)	41.3 (81)
4	More difficult Reading item/question on the computer than paper	16.3 (32)	19.9 (39)	18.4 (36)	18.9 (37)	26.5 (52)
5	Answer questions on computer easier than separate sheet	20.4 (40)	16.8 (33)	17.3 (34)	23.5 (46)	21.9 (43)
6	More anxious on computer than paper	20.9 (41)	27.6 (54)	23.5 (46)	14.3 (28)	13.3 (26)
7	Concern about experience interfered with performance on the computer test	25.5 (50)	21.9 (43)	25.0 (49)	11.7 (23)	15.8 (31)
8	Prefer to take a computer rather than paper test in the future	19.9 (39)	15.3 (30)	27.6 (54)	18.4 (36)	18.9 (37)
9	Computerized tests require too much experience with computers	8.2 (16)	20.4 (40)	19.9 (39)	19.8 (38)	32.1 (63)
10	Wish computerized tests did not bother me so much	7.7 (15)	10.7 (21)	18.4 (36)	17.3 (34)	44.9 (88)
11	Thinking about pressing the wrong key interfered with my performance	19.9 (39)	19.9 (39)	20.9 (41)	20.4 (40)	18.9 (37)
12	Easier to check my responses on paper rather than computer test	13.3 (26)	20.4 (40)	14.8 (29)	19.9 (39)	31.6 (62)
13	Felt more comfortable completing the test on paper than computer	10.2 (20)	14.8 (29)	21.4 (42)	22.4 (44)	31.1 (61)

Table 11. The Frequency of Students Responses in ATCAS

Appendix E. Staff Attitudes towards CAA

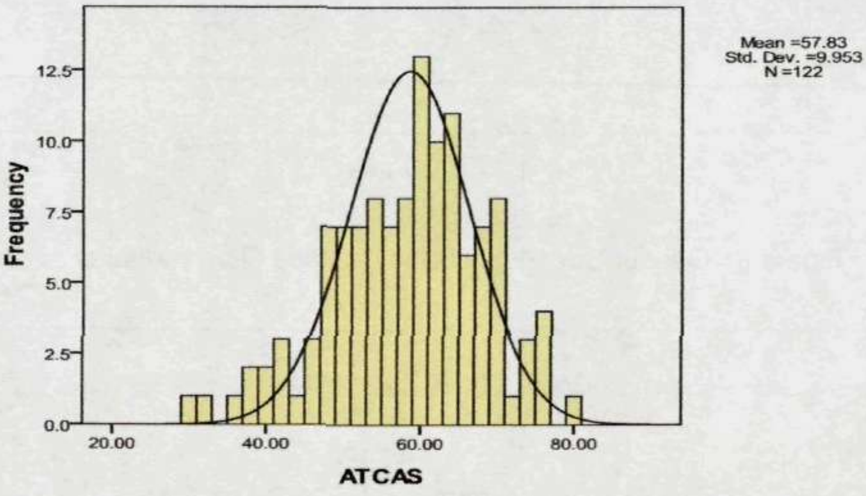


Figure 1. Histograms of Staff Attitudes Questionnaire

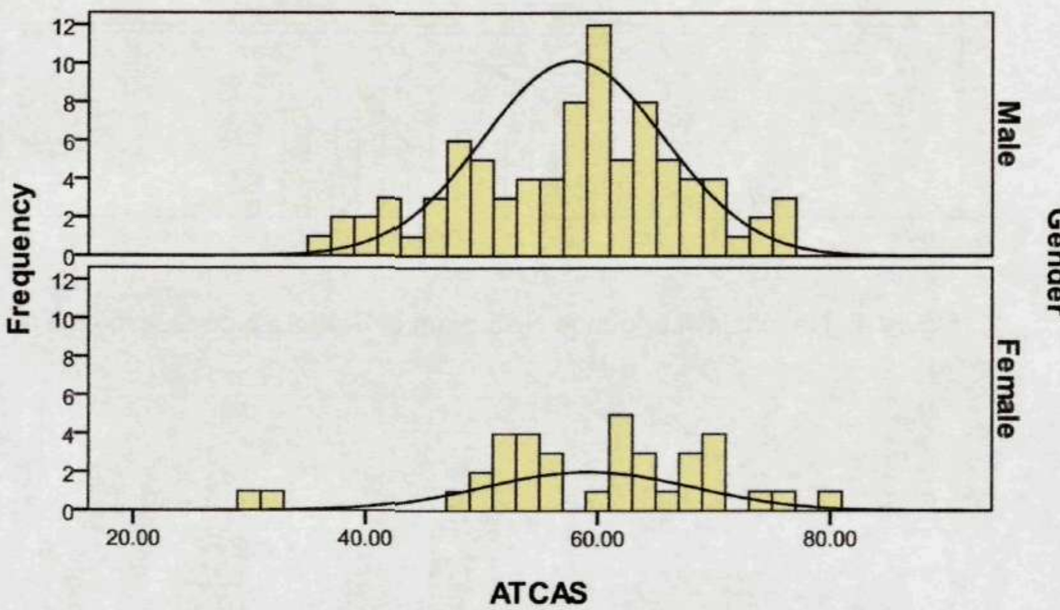


Figure 2. Gender Histogram of Staff Attitudes Questionnaire

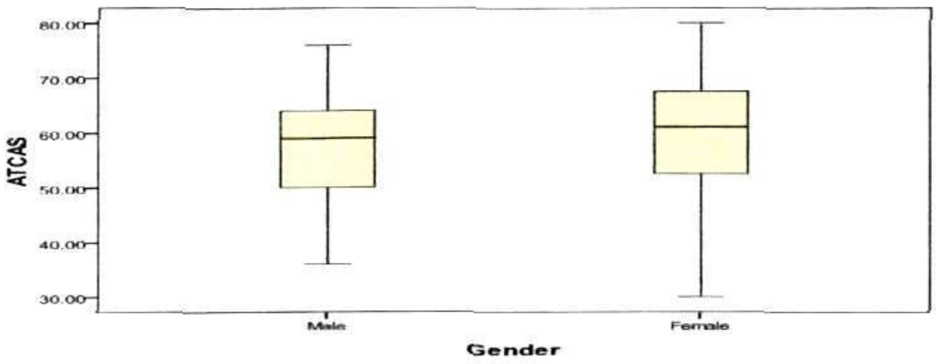


Figure 3. Gender Box Plots of Staff Attitude Questionnaire

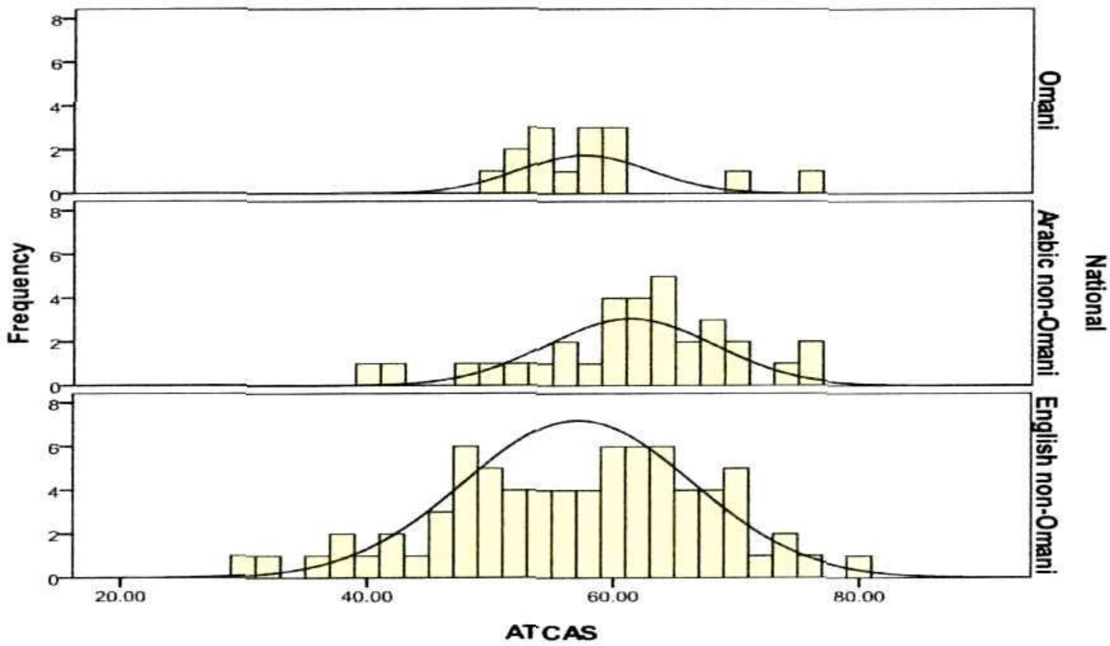


Figure 4. Nationality/Language Histogram of Staff Attitude Questionnaire

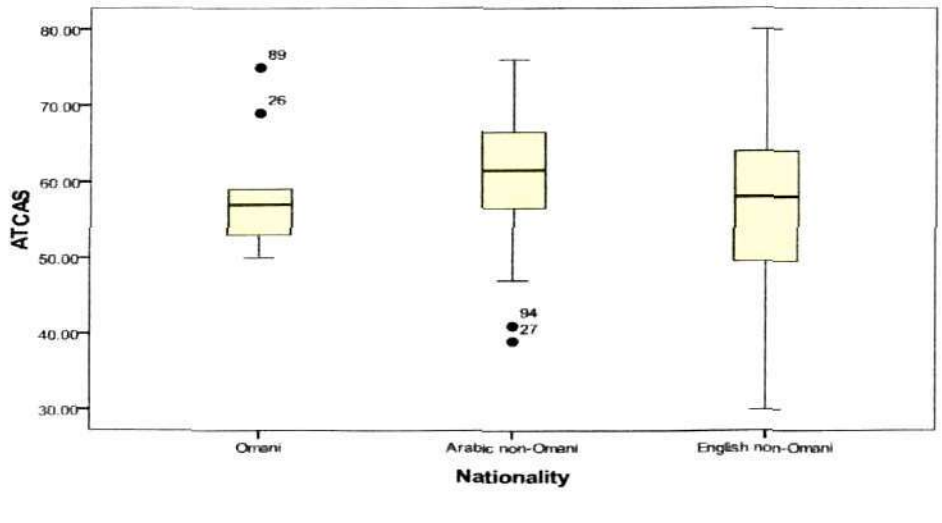


Figure 5. Nationality/Language Box Plots of Staff Attitude Questionnaire

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.852
Bartlett's Test of Sphericity	Approx. Chi-Square	934.695
	Df	153
	Sig.	.000

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	6.308	35.042	35.042	6.308	35.042	35.042
2	1.926	10.700	45.741			
3	1.464	8.131	53.873			
4	1.223	6.797	60.670			
5	1.100	6.110	66.780			
6	.990	5.501	72.281			
7	.810	4.501	76.782			
8	.688	3.822	80.604			
9	.561	3.118	83.722			
10	.499	2.775	86.497			
11	.413	2.296	88.793			
12	.362	2.012	90.805			
13	.346	1.925	92.730			
14	.305	1.693	94.423			
15	.294	1.635	96.058			
16	.273	1.516	97.574			
17	.232	1.290	98.864			
18	.205	1.136	100.000			

Extraction Method: Principal Component Analysis.

Component Matrix[a]

	Component
	1
i1	375
i2	802
i3	384
i4	.651
i5	.779
i6	623
i7	.550
i8	.331
i9	295
i10	710
i11	.447
i12	308
i13	362
i14	.526
i15	.738
i16	.592
i17	.836
i18	811

Extraction Method: Principal Component Analysis

a. 1 components extracted.

Table 1. Factor Analysis for Staff Attitudes Towards CAA Questionnaire

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.882	.878	17

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
i1	57.0924	108.746	.306	.528	.882
i2	58.4118	96.549	.725	.637	.867
i3	57.2437	107.033	.347	.562	.881
i4	58.5714	98.620	.558	.526	.874
i5	58.5882	96.007	.721	.612	.867
i6	58.4286	100.772	.562	.556	.874
i7	58.5630	103.248	.443	.533	.878
i8	57.3529	106.840	.288	.352	.883
i10	58.4202	95.991	.655	.532	.869
i11	58.3529	102.874	.394	.391	.880
i12	57.9580	104.854	.288	.202	.885
i13	58.3697	105.845	.322	.330	.882
i14	58.3866	102.290	.464	.474	.877
i15	58.5714	97.891	.648	.585	.870
i16	57.8655	100.761	.488	.482	.876
i17	58.7815	95.409	.782	.705	.864
i18	58.2185	95.037	.738	.683	.866

Table 2. Internal Consistency of Staff Attitudes Toward CAA Questionnaire

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means		
		F	Sig	t	Df	Sig (2-tailed)
total18	Equal variances assumed	266	607	-.901-	120	.370
	Equal variances not assumed			-864-	60 131	391

Table 3 Gender T-Test Result on Staff Attitudes Towards CAA Questionnaire

Test of Homogeneity of Variances

Staff Attitudes

Levene Statistic	df1	df2	Sig
4.159	2	119	.018

ANOVA

Staff Attitudes

	Sum of Squares	df	Mean Square	F	Sig
Between Groups	420 746	2	210 373	2 164	119
Within Groups	11566.639	119	97 199		
Total	11987.385	121			

Table 4. ANOVA Test Result for Nationality/language Variable Difference on Staff Attitudes Towards CAA

(1) Strongly Disagree (2) (Disagree 3) Moderate (4) Agree (5) Strongly Agree

	Statements %(n)	1	2	3	4	5
1	find computer very useful	.8(1)	1.6(2)	1.6(2)	18.0(22)	77.9(95)
2	like more exams to be administrated using CAA	5.7(7)	13.1(16)	35.2(43)	28.7(35)	17.2(21)
3	use computer at my work	1.6(2)	1.6(2)	4.1(6)	23(28)	69.7(85)
4	computer test is easier than paper test	7.4(9)	18.9(23)	35.2(43)	18(22)	19.7(24)
5	feel confident delivering tests by computers	5.7(7)	19.7(24)	38.5(47)	18.9(23)	17.2(21)
6	questions will be very clear in the screen	3.3(4)	12.3(15)	42.6(52)	25.4(31)	16.4(20)
7	The instructions will be easy in computer.	5.7(7)	11.5(14)	50.8(62)	.18(22)	13.9(17)
8	use computer in my house	4.1(5)	2.5(3)	3.3(4)	25.4(31)	64.8(79)
9	implementing CAA will face lots of barriers	2.5(3)	10.7(13)	31.1(38)	32.8(40)	22.1(27)
10	Dislike assessing student by machines.	20.5(25)	30.3(37)	25.4(31)	14.8(18)	.9(11)
11	computerized test can only assess low cognitive ability	18.9(23)	33.6(41)	28.7(35)	9.8(12)	7.4(9)
12	hesitate to use computer for fear of losing all my work	36.9(45)	32.8(40)	14.8(18)	9.8(12)	5.7(7)
13	using CAA in the second language is difficult	13.1(16)	36.9(45)	35.2(43)	11.5(14)	3.3(4)
14	computerized tests take longer to complete then paper test.	15.6(14)	34.4(42)	27.9(34)	16.4(20)	4.1(5)
15	feel comfortable using CAA with my students.	.9(11)	11.5(14)	40.2(49)	26.2(32)	13.1(16)
16	like to attend computerized test training	4.9(6)	7.4(9)	17.2(21)	30.3(37)	40.2(49)
17	Prefer using computerized test than paper test.	.9(11)	19.7(24))	.41(50)	19.7(24)	10.7(13
18	I think CAA will save my time	7.4(9)	8.2(10)	28.7(35)	30.3(37)	25.4(31)

Table 5. The Frequency of Staff Response in the Questionnaire