User Interface Design Features and their effects on Computer Based Educational Assessment: A Cross Cultural Study

A Thesis

Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Debayan Dhar

(09610507)

Under the supervision of

Prof. (Dr.) Pradeep Yammiyavar

Department of Design

Indian Institute of Technology

Guwahati, Assam, India

October, 2013

Department of Design
Indian Institute of Technology Guwahati
Assam – 781039
India

DECLARATION

I hereby declare that the work contained in this thesis entitled "User Interface Design Features and their effects on Computer Based Educational Assessment: A Cross-Cultural Study" is my own work and done under the guidance of Professor Pradeep Yammiyavar, at the Department of Design, Indian Institute of Technology Guwahati, Assam. To the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at IIT Guwahati or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at IIT Guwahati or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is duly acknowledged.'

Place: Guwahati

Date:

Debayan Dhar
PhD Scholar

Department of Design
Indian Institute of Technology Guwahati
Assam – 781039
India

CERTIFICATE

This is to certify that the work contained in this thesis titled "User Interface Design Features and their effects on Computer Based Educational Assessment: A Cross-Cultural Study" submitted by Mr. Debayan Dhar to the Indian Institute of Technology Guwahati for the award of the degree of Doctor of Philosophy has been carried out under my supervision. This work has not been submitted elsewhere for the award of any other degree or diploma.

Place: Guwahati

Date:

Prof. (Dr.) Pradeep Yammiyavar, FDRS
Supervisor
Department of Design
Indian Institute of Technology Guwahati
Assam -781039, India.

Abstract

This thesis investigates human computer interaction issues within the context of ‘examination’ conducted through the medium of computer. By ‘examination’ we mean the ‘assessment’ a student undergoes during or at the end of an academic cycle of learning. In research literature the term ' Test Mode' is found to be used extensively to describe the effects of computer as a medium / tool on the performance outcome of the user (in our case a student under examination). There are evidences from various published empirical studies which suggests that when identical pen- paper based tests and computer based tests are compared they do not produce the same result.

In this thesis it is argued that different formats of presentation of test items through the computer based media can also affect students’ performances. A cross cultural study has been conducted and effect of various features of GUI (Graphical user interface) has been investigated. It is found that there is a differential impact of GUI features on students’ performances across cultures. Therefore, it is argued here that a set of GUI features should be optimized for CBA assessment system such that they affect students’ minimally.

Table of Contents

List of Tables

List of Figures

Abbreviations used in the thesis

CBA	Computer Based Assessment
CBT	Computer Based Test
PBT	Paper Based Test
P&P	Paper and Pencil
NASA TLX	National Aeronautics and Space Administration Task Load Index
SUS	System Usability Scale / Score
PEOU	Perceived Ease of Use
PEU	Perceived Usefulness
SQ	System Quality
RA	Relative Advantage
WCU	Willingness to Continue to Use
PES	Perceived E-learner Satisfaction
IE	Instructional Efficiency
GUI	Graphical User Interface
UCD	User Centered Design
LCD	Learner Centered Design
ICT	Information and Communication Technology

Acknowledgements

First and foremost I want to thank my PhD supervisor Professor Pradeep Yammiyavar for his warm encouragement and thoughtful guidance. It has been an honour to be his PhD student. He has taught me, both consciously and unconsciously, how good experimental research is done. I appreciate all his contributions of time, and ideas to make my Ph.D. experience productive and stimulating. I also thank the other members of my doctoral committee: Professor Debkumar Chakrabarti, Professor Amarendra Kumar Das and Dr. Samit Bhattacharya for their time, interest, and helpful comments.

I am especially grateful to my Fulbright Doctoral program advisor Professor Samuel Gosling, for inspirational discussions, encouragement and insightful support during my Fulbright Program period. I am thankful to Jason Ferrell, Lindsay Graham, Gabriella Harari, Abby Black and Selina Yanez for providing me whole hearted help in data collection and transcription during the study conducted in USA. I would like to thank Justin Davis, Senior Systems Administrator at Liberal Arts ITS, UT Austin for his support and help during the installation of the online experiments at UT server.

I would like to thank Suman Debnath for helping me out with the technicalities during the development of the online experiments. I owe to Sandip Adhikary and Dandeswar Bisoyi for proving whole hearted help in Data collection in India. I would like to express my gratitude to my friend Steve Sullivan who made my stay in the United States a comfortable one.

But above all I would like to express my heartfelt gratitude to my good friend Dr. Susanne Beier not only for her excellent comments but also for listening to me whenever I was excited about a new idea. I am thankful to my friends Thainswemong Choudhury, Aditya Ponnada, Carlotta Ricci, Soumyadeep Chakraborty, Samir Kumar Panda, Sangeeta Pandit, Pitambar Ranadive and Yogesh Deshpande for their support. My thanks are due to staff, faculty and students at DOD who helped me and provided support for the studies. I thank the Fulbright Foundation for providing support for conducting the studies in USA.

Last but not the least I would like to thank my family members for holding my hand during tough times and for motivating me for the studies.

Debayan Dhar

FUNDING ACKNOWLEDGEMENT

The author hereby duly acknowledge the Fulbright Foundation and the United States India Educational Foundation for awarding the 'Fulbright Doctoral Fellowship' and sponsoring the research studies at the Department of Social Psychology, University of Texas at Austin under the supervision of Professor Samuel Gosling. The experimental work in USA, presented in this thesis was possible entirely due this funding and opportunity given by the Fulbright Foundation.

The author also acknowledges the Indian Institute of Technology Guwahati, for appointing him as a Teaching Assistant and supporting the research studies conducted in India by the author under the supervision of Professor Pradeep Yammiyavar.

Place: Guwahati

Date:

Debayan Dhar
PhD Scholar

Chapter 1

Computer based Assessment and Testing

Chapter abstract: *This chapter highlights the research issues that are being dealt with in this thesis. Current issues are introduced and placed in the context of their background. The objectives of the research are reported, and highlighted. The boundaries and scope of the thesis is laid out along with definitions of terminology. Summaries of all the Chapters are outlined.*

1.1 Introduction

The use of Computers in education has been one of the fastest absorption of computing technology by a sector that involves millions of human users in the form of students as well as teachers. One estimate indicates that up to 10.8 million units of computers were sold in the year 2011-2012 in knowledge hungry country like India (Singh, 2012). The market size of Information and communication technology (ICT) in India in the year 2012 was INR 144 billion and it is projected that the market size is set to reach INR 384 billion in 2014 ("India: ICT led enterprise", 2012). ICT in the educational sector has witnessed a massive change in investments and many institutes in India have begun delivering educational materials through the ICT media. In the coming decade with increasing penetration of internet into the rural hinterland, the education sector in India will witness a massive change.

Computers are used as an instrument for conducting experiments and as a tool for research but apart from this there has been wide scale deployment of the computer as a device for storing and consuming processed knowledge. The term *edutainment* or *infotainment* is often used in popular literature to describe the educational value of the computer. It is but natural that the computer has also emerged as the most suitable medium for taking or conducting an examination. The use of computers in the education sector continues to increase exponentially in administration, in research and development and more important as a mass learning and teaching tool.

Many Indian educational institutions now conduct computer based entrance examinations such as CAT (Common Admission Test – conducted by the Indian Institute of Managements), BITSAT (Birla Institute of Technology and Sciences Aptitude Test), etc. Many Indian students seeking admission beyond Indian shores take online computer based exams such as GRE (Graduate Records Examination), TOEFL (Test of English as a Foreign Language) etc. The advantages they offer are well-known and obvious. *Computer Based Testing* (CBT) offers standardization of test

administration. In case of a large test population size, CBT helps test developers to set the same test conditions for all examinees. It improves various aspects of test security by storing questions and responses in encrypted databases. It enables testers to create randomized questions and answers from vast question pools. On the examinees' side, they are able to receive greater measurement efficiency and the possibility to take tests at any time and place convenient to them. In fact, CBT which is also referred to as *Computer Based Assessment* (CBA) has overcome the temporal and spatial constraints posed by traditional format of educational assessment and as such has become an effective tool of learning and teaching.

On the other hand, there are some disadvantages that students' have to become aware of before opting for computer-based testing which led many research scholars to suggest systematic studies to check equivalency and comparability of paper-based tests with computer-based tests (Parshall et al., 2002). For example, students need some degree of computer literacy in order to avoid the mode effect on computer-based testing (Alderson, 2000). According to instructional design dogma, paper based versus computer mediated instructional components should produce exactly equivalent results if the content and the cognitive activities of the two tests are identical (Clark 1994). In most test mode effect studies, computer based and paper based versions are nearly identical and cognitive activity required to answer a test item on paper or computer are almost same, yet significant differences are regularly observed (Bunderson, Inouye & Olsen, 1989). Literature on test mode effect studies are inconclusive and arguments still exist regarding which delivery media should be crowned as the most efficient one for delivering the assessment content. To address the issues raised in research literature investigations need to be carried out to study the effect of various computer based tests on the psychology of the examinee (the test taker) more closely in order to make the CBA format more empathic towards the examinees.

In this thesis we intend to understand the human computer interaction issues within the context of 'Examination' conducted through the medium of computer. By 'examination' we mean the 'assessment' a student undergoes at the end of an academic cycle of learning. In research literature the term 'Test Mode' has been found to be used extensively to describe the effects of delivery media / tool on the performance outcome of the user (in our case a student under examination). There are evidences from various empirical investigations published in literature which suggests that when identical *pen- paper based tests* and *computer based tests* are compared they do not produce the same result. The mode of taking the test (example: on paper or on computers) has been argued to have an effect on the end results of the test for the test giver. Such findings are referred to as *"test mode effect"*. It is known widely that in a traditional paper based test situation, individual learner characteristics of the students and their interaction with the

presentation format in traditional (pen and paper) testing environment affects examinees' success in a test situation. There are many variables that act on performance efficiency of students and on which the outcome of the test depends. Can it therefore be assumed that the same variables interact and affect the students' performance in a computer based test format? Is it simplistic to assume that the computer has only replaced the 'paper' and as such this need not have much bearing on the examinee or the performance? There are two aspects to performance - physical and mental. This thesis work investigates the mental aspect which is often termed as the 'cognitive aspect' of the students from the perspective of an assessment frame work.

Literature in 'test mode effect' studies points out to various contradictory findings and questions that still remain unanswered. The state of the art review conducted (chapter 2) highlights certain gaps that were investigated in the present thesis work. The literature survey points out:

i. Inconclusive research from the perspective of identifying the effect of GUIs (Graphical User Interfaces) on the individual characteristics of the Examinees.

ii. The literature highlights how individual characteristics affect the interaction between the examinees and the computer based tests and argues that individual characteristics can be a predictor of test mode effect. But past research literature did not investigate the critical individual characteristics' that might affect students' performance in an assessment environment.

It is therefore posited here in this thesis that the *effect* of the *delivery media* which in this case is a computer based one, on the student's individual characteristics' should be investigated to address the issue of equivalency and comparability of testing formats in an assessment environment. By presentation format we mean the manner, by which test questions are presented on a computer screen in terms of their sequences, frequency, graphical and typographical characteristics and overall layout of the question paper.

1.2 Objectives of the Research

The broad objectives that set the framework for refinement of the Research problems (which will be dealt in detail in Chapter 2 and in the subsequent chapters) are stated thus:

To study & ascertain by observation, systematic analysis and experimentation

i. The role of presentation format and how it relates to individual learner characteristics. In other words is the effect of computer (as a medium of interaction) the same on all individuals or does it differ with the students' individual learning characteristics and traits.

ii. To investigate how presentation format might interact with individual learner characteristics and the extent to which it facilitates examinee performance.

iii. To ascertain if there is a need to evolve a different instructional paradigm in the case of a computer as different from the instruction style while using paper based testing.

iv. To find out if the examinee's cultural context has an influence on their performance. In other words is there a difference between the Indian students learning and interacting context as compared with students from another country or culture.

v. To evolve heuristics that will be useful for a GUI designer in his / her work of content creation & information architecture while designing formats & Interfaces for assessment applications.

1.2.1 Rationale for the problem chosen to be investigated

It is the contradicting findings and differences in "test mode effect" studies that have resulted in the motivation to carry out the research which is being reported in this thesis. The wide spread use and popularity of large scale computer based on-line entrance examinations involving millions of students in India is serious enough to be improved so as to minimise the interacting effect of the presentation format that might induce additional work load in addition to the normal workload of an academic examination on the students and may affect their test performance.

Presentation of the assessment content in the traditional mode (pen and paper format) provides examiners with limited scope for presentation of the test items. As a result, almost every pen and paper based test format remains identical in terms of their presentation format. Whereas computer based technologies provide freedom to the educational test designers who can present the test items innovatively and in such a way wherein no two CBA formats are identical in terms of their presentation format. One important issue with the design of CBA systems is that often User-Centered design (UCD) framework is widely employed and used for design and development of the screen interfaces of a CBA system.

Hsi and Soloway (1998) suggests that though focuses on reducing users' cognitive load, makes the system - easy to learn and reduces time-on task, using this framework alone cannot address the complexity of human learning. While the "users" in an UCD approach is considered to be knowledgeable and motivated about their tasks a student cannot be considered so by default especially during stressful exam conditions. They lack motivation and domain knowledge. Hence

designing for such users who have unique needs, strategies need to be employed to address these specific needs in order to satisfy users who use such systems. The success of e-learning system does not only rely on a usable and accessible system but also depends on how efficiently the system aids the learning process of the users to achieve their desired learning goals. In the context of assessment (test or examination) it can be said that the assessment system should assist the students in solving the assessment tasks and help them in achieving the assessment objectives. As mentioned by Hsi and Soloway (1998) the goal of Learner-Centered design approach is to design software that "make people more effective learners," it is based on the knowledge of the users and their different characteristics: how learners prefer to learn, how they are learning the information, under what pressures the learners operate in their day-to-day life, their motivation or incentive to engage in online learning, what constraints they face, what special accommodations they need, how they feel comfortable while using the online applications, what experience they have with e-learning applications and so on.

Literature in 'test mode' effect studies that investigated equivalency of presentation modes has hardly considered the unique learner characteristics that are affected by the new mode of presentation. Few studies that attempted to address the issue have highlighted that computer literacy has an effect on examinees' test scores. Computer literacy of such examinees actually means that their prior experience, acquaintance and engagement with the computers have prepared them to understand which parameters they need to consider to make an easy transition from a paper based test to a computer based test so that the CBA mode do not affect their efficiency. During these interactions with new technologies students try to align themselves with the new mode of interaction and create a mental model of their experiences. More investigations are required to address the equivalency of test scores across modes as test mode effect still exists. Apart from computer literacy, the unique learner characteristics that might get affected due to the new mode of presentation are required to be investigated. In order to understand these kinds of interactions and learner experiences we need to understand the learner activities in an assessment environment and how they react to emotion and affect especially in the context of an assessment (examination) using a computer. This thesis reports the systematic study, experimental methodology and the findings that will help address the objectives.

1.3 Computer –Based Assessment & definitions - setting frameworks

Assessment is often described as the activities that the tutor undertakes to help learners learn and quantify the learning progress and outcomes (Black & William, 1998). It is vital for learning as it assesses the learning outcomes of a student which is a cognitive process of acquiring

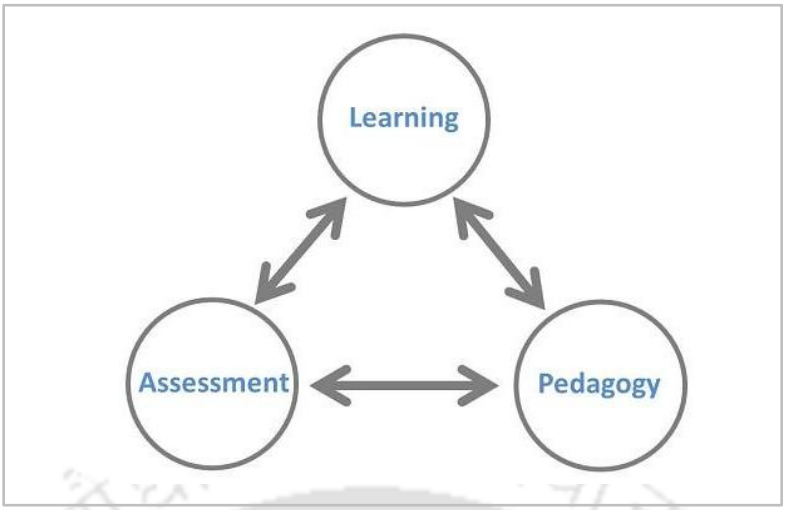


Figure 1.1 Intimate association of Learning, Assessment and Pedagogy; adopted from Pellegrino, Chudowski and Glaser (2001).

skills or knowledge. It measures and documents the knowledge, skills, and attitudes of an individual learner or a learning community (e.g., class, course, or workshop), or an educational institution (Amelung, Krieger & Rosner, 2011) and paves the way for further learning. Assessment evaluates student learning through an assessment method (e.g., assignments, examinations, laboratory work and portfolios).

Assessment methods are crucial in the fulfillment of the objectives of the course, i.e., the expected study outcomes. The choice of the assessment methods may be linked to the overall objectives of the specific study program and may include the development of the analytical disciplinary skills (e.g., critical evaluation or problem solving) and / or support the development of generic professional competencies (e.g., particular team skills, group work). Computers and internet provides new possibilities for delivering learning content (e-learning) and examinations. Computer based assessment is one of the domains of e-learning which involves the use of electronic technology (computers and internet) for delivering the assessment content wherever the student is. Due to increase in the use of e-learning, computer based assessments (CBA) are often referred to as e-assessments, online exams, computer based tests etc. However, all refer to the same concept of electronic form of delivery of the assessment content.

Pellegrino, Chudowski and Glaser (2001) highlight the intimate association of teaching (pedagogy), learning and assessment (Figure 1.1). Students' performance in assessment plays a significant role in shaping the teaching style or designing the curriculum. While many researchers have studied the interrelatedness of the three - Learning, Pedagogy and Assessment with respect to

traditional paper based systems; in this thesis deeper understanding of the influence of the computer on to the examinee is attempted to be addressed.

According to Robitaille, Schmidt, Raizen, McKnight, Britton, and Nicol (1993), assessment provides a clear and definitive guidance of what is to be learned than any verbal description and hence is a better basis for curriculum planning at classroom level and the most potent driver of classroom practice (Ridgway, McCusker & Pead, 2004). Assessment affects students learning as well. It analyses and evaluates students' skills to ensure that students attain the desired learning objectives and gain confidence in their abilities. The basic functions and objectives of educational assessments as highlighted by Mariyana Nikolova (2012) are stated below:

i. Evaluation (feedback) - to measure the results of the learning process.
ii. Diagnostic – to detect individual educational problems.
iii. Forming – to direct and manage the learning process.
iv. Motivation - to stimulate and to motivate.
v. Perfection – to improve the students' knowledge.
vi. Providing – to collect data and to process statistical evaluation.

The objectives of assessment as highlighted by Bresciani (2002) are as follows:

i. To reinforce or emphasize the mission of the unit. Improve programs and/or performance (formative).
ii. Compare a program's quality or value to the program's previously defined principles (summative).
iii. Inform planning.
iv. Inform decision making.
v. Inform policy discussions at the local, state, regional, and national level.
vi. Evaluate programs, not personnel.
vii. Assist in the request for additional funds from the University and external community.
viii. Assist in the re-allocation of resources.
ix. Assist in meeting accreditation requirements, models of best practices, and national benchmarks.
x. Celebrate successes.
xi. Manage expectations.
xii. Reflect on the attitudes and approach we take in improving teaching and learning.

xiii. Build a learning culture grounded on the values of accountability and provides an opportunity for continuous improvement.

The assessment process incorporates a number of tasks required to complete the assessment cycle. The normal life cycle of assessment starts with the preparation of the assessment component, then administering the assessment component to the students, tracking the assessment progress, evaluating and marking the response sheets of the students, moderation of assessment results, collation and storage of assessment results, giving feedback to students, post-assessment session evaluation- comparison of results for completed assessments (Prakash & Saini, 2012). Traditional pen and paper based assessments require a high degree of coordination especially if the class size and number of educators teaching the course are large. Therefore, it becomes a burden for the assessment administrators to effectively administer the assessment cycle. Moreover, a large class size significantly increases the load on the educators as a result of which efficiency of the assessment process decreases significantly. As student numbers have increased many universities have introduced computer-based assessments as a mechanism for providing immediate feedback to students and to reduce load on over-stretched staff (Bull, 1993; Lyell & McNamara, 2000).

Moreover according to Pellegrino, Chudowsky and Glaser (2001) the traditional pen and paper mode of delivery of the assessment content fails to capture the breadth and richness of knowledge and cognition. This is mainly because of the fact that the traditional assessment practices generally focus on assessing whether a student has acquired the content knowledge but they often fail to assess the learning process and higher-order thinking skills (Baek, 1994; Bahr & Bahr, 1997). Computer based assessments are anticipated to outclass the traditional mode in this regard by incorporating on going and multiple assessment strategies into the learning process (Lin & Dwyer, 2006).

Joint Information Systems Committee (JISC) defined computer based assessment (e-assessment) as, "the end-to-end assessment processes where Information and Communication Technology (ICT) is used for the presentation of assessment activity and the recording of responses" [JISC]. Computer based assessment (CBA) when used as component of e-learning or as an independent entity, helps students to evaluate their strengths and weaknesses (Joosten-ten Brinke, van Bruggen, et al. 2007; Kaklauskas, Zavadskas, et al. 2010).Computer based assessments (CBAs) are developed in order to automate the assessment process (Charman & Elmes, 1998; Economides & Roupas, 2007; Fluck, Pullen, & Harper, 2009; Gvozdenko & Chambers, 2007). During the initial days the use of CBAs was restricted to simple text-based question and answer sessions written by programmers. As information and communication technology developed authoring systems came into the foray and CBAs became more user-friendly and gained wide

access in academics (Whiting, 1985). Numerous free and easy to use CBA products are available online today (Derbyshire, 1999). Due to the advance in ICT technologies today questions in a CBA need not be only text-based but a variety of graphics and multimedia can also be incorporated. Similarly, students can now answer question through a variety of input mechanism which include multiple choice selection, graphical hotspot clicking, text, numerical and mathematical answers. CBA are now used extensively in various educational institutions during entrance tests, final exams, mid-unit tests and diagnostic assessments.

1.3.1 History of Computer based Assessment

The first computer based assessment system that was developed for supporting marking and grading of student solutions dates back to as early as 1960 (Saikkonen, Malmi, & Korhonen, 2001). Later the introduction of microcomputers in the 1970s led to the widespread development of computer based versions of the classical paper-and-pencil and electromechanical tests of human capabilities. Gradually the computer based human performance tests evolved into more advanced computer-based assessment batteries (CABs).

The computer-based assessment batteries (CABs) provided automated assessment of a wide range of skills and abilities (Gilliland & Schlegel, 1993; Kane & Kay, 1992; Perez, Masline, Ramsey, & Urban, 1987; Schlegel, Hockey, & Gilliland, 2003; Shingledecker, 1984). The CABs when used demonstrated their superior effectiveness in comparison to the classical pen and paper based tests in human performance research and real-world applications including personnel screening and selection, general medical applications in the areas of neuropsychology and disease assessment, and a vast array of educational, industrial, and business applications.

1.3.2 Benefits of CBAs

The extensive use of computer based assessments can be attributed to the innumerable advantages it offers over the traditional pen and paper based assessments. According to Bugbee, 1996; Drasgow and Olsen-Buchanan, 1999; Mazzeo and Harvey, 1988; Mead and Drasgow, 1993; Parshall, Spray, Kalohn, and Davey, 2002; Tseng, Macleod, and Wright, 1997; Bennett, 1998; Chatzopoulou and Economides, 2010; Scalise and Gifford, 2006 the benefits of CBA are as follows:

i. Increased test security.
ii. Reduced running cost.
iii. Instantaneous score reporting.

iv. Precise measurement of test scores and response times.
v. Automatic record keeping for item analysis.
vi. Offers innovations in testing.
vii. Flexibility in time and place.
viii. Usability of a greater variety of media and test types.

In light of these advantages, many employers, psychologists, educators and researchers have converted conventional or paper-and-pencil (P&P) tests to computer-based test (CBA) formats (McDonald, 2002; Neil, 1996; Segall & Moreno, 1999; Van de Vijver & Harsveld, 1994). The disadvantages of CBAs are listed in detail in Appendix 4, page 238.

1.3.3 Types of CBAs

The two major categories of computer based assessments often mentioned in literature and in practice are:

i. Formative Assessments.
ii. Summative Assessments.

Summative computer based assessments, allows the teacher to evaluate a student's abilities at a given point, as the basis for a decision about pass or failure. It also helps the students to evaluate their effectiveness in learning. On the other hand, formative assessments help students in reaching their targets through appropriate feedbacks. It is defined as judgment combined with immediate feedback – it enables the learner to identify and close possible 'gaps' between the actual level of work and the required standard (Birenbaum, 1996; Turner & Gibbs, 2010; Cantillon, Irish & Sales, 2004; Gikandi, Morrow, & Davis, 2011; Taras, 2005). According to Thelwall (2000), CBA can be further classified based on different contexts, and the different functions it performs. Table 1 highlights the further classification of Summative and formative CBAs. An assessment can be formative in two different ways:

i. By providing feedback on performance and
ii. By the skills gained while performing the task.

The taxonomies of various CBAs provided in table 1.1 contain parts of both formative and summative assessments but in varying degrees. Grading tests and diagnostic tests primarily rely on feedback on the other hand exercises aims to improve skills. Assessments which are not purely formative require more security and hence are complex than the classical pen and paper based tests (Zakrzewski & Bull, 1998). Traditionally written tests were stored in a sealed envelope, but in a

Table 1.1

A Taxonomy of applications of CBA adopted from Thelwall (2000).

Area	Type	Description
Summative	Exam	An assessment solely for grading purposes such as exam at the end of a unit of study.
Formative/Summative	Grading test	An assessment for grading but which also provides feedback intended to direct future studies. Includes a mid-unit small test, or weekly problem sets.
Formative	Open access test	A grading test that doubles as a set of exercises because students are allowed to practice before sitting the test.
Formative	Self-test	An assessment designed to give feedback to a student on their progress with a section of a unit of study (or test).
Formative	Exercises	A problem set designed to consolidate learning on a section of a unit of study.
Formative	Programmed learning tool	A linear CAL package-based upon a question and answer session.
Formative	CAL Quiz	A marked exercise integrated into a CAL package, for example a multiple choice question presented after a slide containing new information.
Formative	Adaptive CAL quiz	A marked exercise integrated into a CAL package used to test the student but also used to adapt the teaching of the package to student weaknesses.
Formative	Diagnostic test	An assessment of prior learning taken before a unit of study.

CBA the test content is usually stored on a network in advance of the test and remains protected from student access (Thelwall, 2000).

1.3.4 Types of test administration

The assessment content is administered via the computer based assessment system mainly in two broad categories, they are – power and speed tests. Pure power tests assess an individual's ability in a specific content area with no regard for how quickly items are answered. Whereas speeded tests are designed to measure processing speed and thus stringent time limitations are imposed. Often educators administer timed-power test, which is a type of power test with some limitation on time and thus depends on power and speed in varying proportions (Anastasi & Urbina, 1997).

1.3.5 Types of delivery of CBA systems

Computer based assessment is delivered to students through two basic formats, the web-based delivery format and as a stand-alone desktop application. In the web based delivery format, the assessment content is accessed from a computer through the internet connection by the students. In the web based format the transfer of data prior to and after the assessment session is conducted via the internet. In case of a stand-alone desktop application the encrypted test is downloaded onto computers in the appropriate test centre/s at a designated date and time and then released when the candidate arrives to take the test.

1.3.6 Assessment content and CBA delivery systems

The aim of any assessment component is to test the learning outcome. It is prepared with the intention of addressing a particular learning outcome / area or related areas of knowledge in the course. This can be termed as mapping of a learning outcome. Performance of the students in terms of test results can be collated to review the knowledge acquisition of the learner and mastery of core learning outcomes as outlined in the course profile. Figure 1.2 (next page) highlights the CBA workflow model (adopted from Prakash & Saini, 2012). The instructor prepares the assessment component to test the learning outcome of the students. Learning outcomes are perceived as directly observable performance improvement in assessment and evaluation. Once the CBA is administered the test scores are obtained. Test scores are indicators of a student's learning effectiveness and his knowledge gained from classroom interaction (Gagné, 1985). Piccoli, Ahmad and Ives in 2001 highlighted learning outcomes as one of the indicators of the effectiveness of e-learning. Though it is desirable that all learners acquire the same learning outcomes as same level or at least a comparative level, it has been observed that this is not the case as most learners acquire different learning outcomes and at various levels (Marton, Booth, & Booth, 1997).The entire CBA system is managed by two important components: the assessment engine and the item bank. The assessment engine comprises the hardware and software required to create and deliver an actual test. Most CBA engines run on standard hardware so the key characteristic of the CBA system is the software's functionality. The software does not include the questions themselves; these are provided by an item bank. Once created, the engine uses the item bank to generate a test. For a successful and efficient CBA system the item bank and the data transfer should be secured so that security threats can be minimized. Test developers around the world have begun to take measures to deal with some

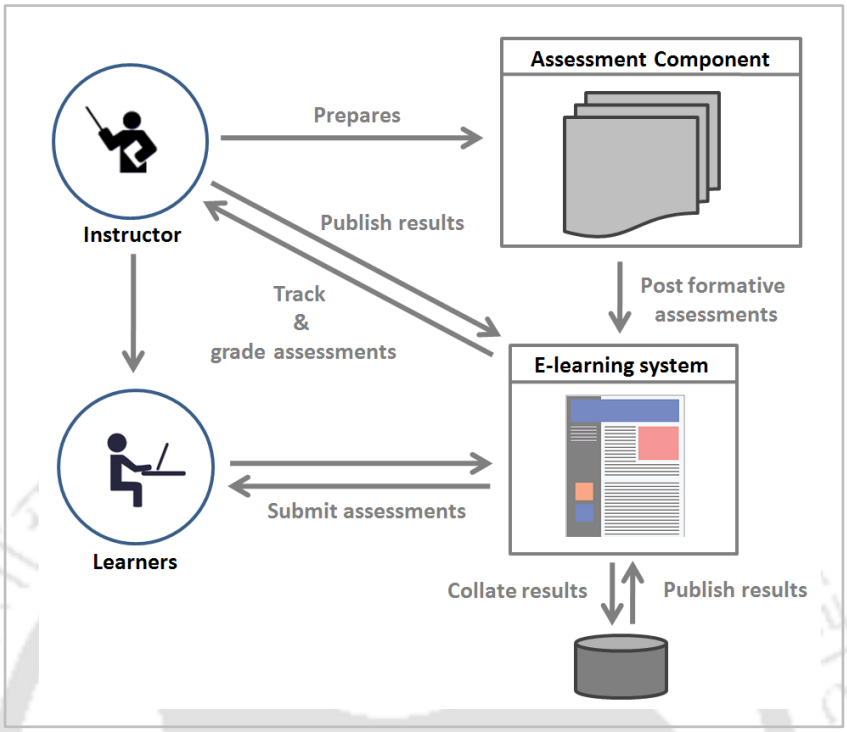


Figure 1.2 CBA in an E-Learning environment adopted from Prakash and Saini (2012).

very real threats to the integrity of their testing programs posed by item-pool exposure collaboration and other forms of cheating.

High stake CBAs (e.g. entrance into graduate school, scholarships, a coveted course placement, a job, a license, a professional certificate), always face high security issues due to some group of cheaters who may employ well-thought-out strategies that may provide them with any possible advantage, however slight that may be.

There are four general methods for dealing with risks in high-stakes CBT:

i. Using randomization schemes to scramble items and other "test units" as much as possible.
ii. Increasing the size of active item pools.
iii. Rotating item pools (intact or partially) over time.
iv. Specifically controlling item exposures as part of the computerized test assembly process (e.g., using item-level exposure control mechanisms) (Hetter & Sympson, 1997; Sympson & Hetter, 1985; Stocking & Lewis, 1995, 1998; Revuela & Ponsoda, 1998).

In high-stake CBA environments the best test-delivery models are those that minimize the greatest number of risks and simultaneously reduce the magnitude of specific security risks, all without requiring extensive sacrifices or trade-offs elsewhere and without substantially adding to overall costs (Luecht & Sireci, 2011).

1.4 Issues related to CBA systems

Two pertinent issues relating to the use of CBAs are the equivalency of computer-generated scores and corresponding pen & paper scores, and the validity of test score interpretations (Clauser, Kane, & Swanson, 2002). Equivalency is concerned with numeric score comparability, while the validity of CBT interpretation is concerned with the underlying constructs that the test purports to measure (Association of Test Publishers, 2000; Van de Vijver & Harsveld, 1994). Attempts to establish construct equivalence provide valuable evidence on whether the attributes measured by computerized assessment methods are similar to those attributes assessed through conventional methods. Studies that investigated the equivalency of CBA test scores with pen and paper based test scores are often referred to as test mode effect studies. Situations wherein traditional tests and CBAs are used simultaneously and test scores from both the format are compared, then, if the testing formats are not equivalent valid test score interpretation would provide educators with incorrect performance measurement of the student population. The next chapter highlights the state of the art of studies conducted in test mode effect research and reports the research problem that guided the investigation of the thesis work reported here.

1.5 Scope of the thesis:

The work in this thesis is in the area of Human Computer Interaction design. It spreads across disciplines such as Psychology, Educational Technology, and Pedagogy & Human Centered Design. Though the current research work reviews literature in "test mode effect" studies, but the research investigation argues that in order to reduce the performance difference across the two modes, first effects of various presentation formats in a CBA condition (CBAS provide instructional designers with the innovative use of testing formats) should be investigated and thereafter an optimized set of presentation format for the CBA should be decided which can then probably be compared with the pen and paper based format for investigating test mode effect.

The **presentation format effect** is the focal point around which research questions have been formulated, hypotheses postulated, experiments conducted, results analysed and conclusions inferred. In this thesis it is argued that test scores which has been cited extensively in literature is not a sufficient indicator of or a measure of 'test mode effect' that may happen due to the very

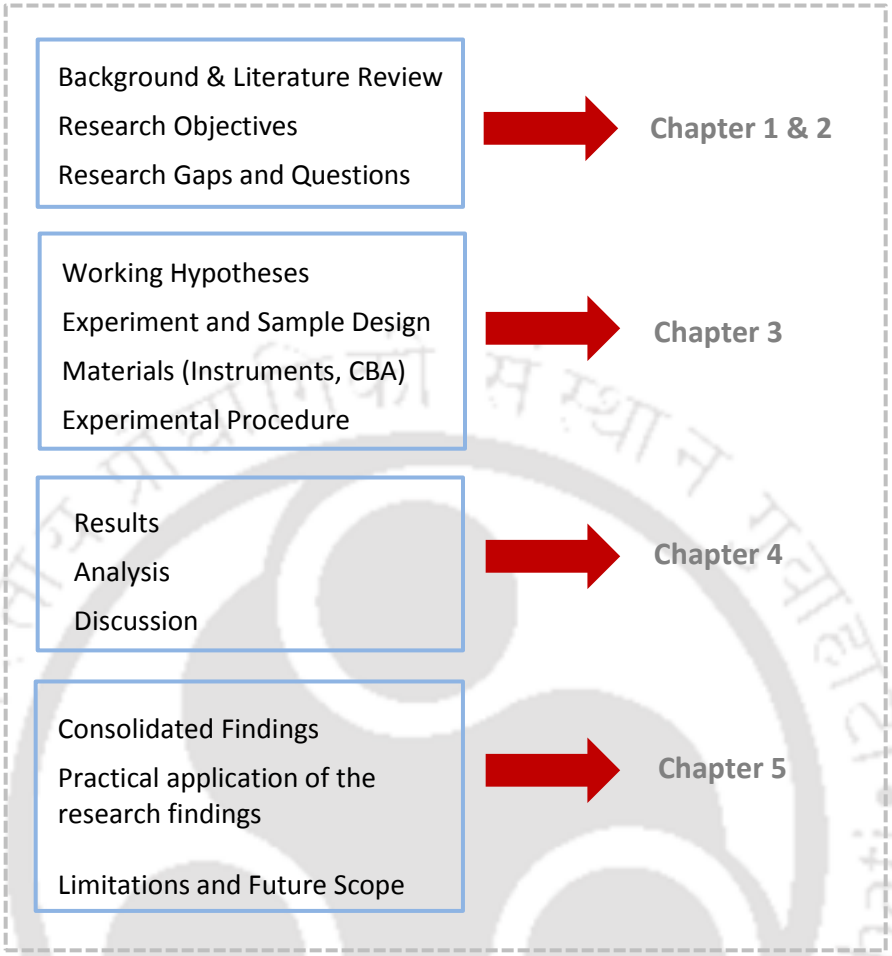


Figure 1.3 Broad research methodology followed in the thesis

visual layout of the computer interface. Other unique learner characteristics should also be considered which may affect the students in a computer based assessment environment. Moreover, since computerized delivery of the assessment contents provides test designers the flexibility to design innovative item types and presentation format it is argued here that various presentation formats might affect students' performance in a computer based assessment environment. Therefore it is required that investigations are carried out to optimize presentation formats that don't interfere with the students' individual characteristics and their performance. In other words , we intend to compare between 'Presentation Formats" (specifically the interactive components of GUI's) of CBA's to investigate and answer the question whether Presentation Formats can be held responsible for variation in the users performance in an examination - where test content complexity as well as the intellectual capacity of the user is assumed constant throughout the various experimental conditions. The broad methodology in which this thesis unfolds is as shown in the Figure 1.3.

1.6 Chapter Summaries

Chapter 1: The Research issues being dealt with in this thesis are introduced and placed in the context of their disciplinary background spanning Human Computer Interaction, Educational Psychology, Pedagogy and Instructional Design. The objectives of the thesis are highlighted. Historical work published on the topic has been cited .The boundaries and scope of the thesis is laid out along with definitions of terminology. Summaries of all the Chapters are outlined.

Chapter 2: A State of the art review on literature in 'test mode effect' studies has been presented. Specifically, past studies on the effect of various interface features in a computer based assessment environment on students have been discussed in detail. The impact of the past investigations on current CBA trend has been reported. Issues with past investigations have been reported and research gaps discussed. Rationale for the stated research questions has been highlighted.

Chapter 3: This chapter presents the working hypotheses. The independent measures and the dependent measures have been listed. Materials (Instruments) used to collect data have been cited and their measure of reliability reported. Sampling frame has been highlighted and the sampling procedure to collect data reported. The treatments have been elaborately discussed and the CBA system designed for the experiments have been described. Procedure adopted for data collection is reported.

Chapter 4: This chapter reports the statistical analysis of the data collected in the experiments. Cross cultural analysis of the data and the effect of the interface features on the students have been reported. Detailed discussion of the results from the light of the literature has been reported. It has been discussed how the current results of the investigations throws light and provides direction to the 'test mode effect' studies.

Chapter 5: Consolidated findings of the investigation have been reported and the implications of the findings for designers of GUIs have been highlighted. Future scope and the limitations of the current study have been presented.

1.7 Chapter 1: Summary

The research issues investigated in this thesis has been presented in this chapter. CBA and issues related to it has been highlighted. The aim of the current research work has been reported and the research objectives are placed in the context of their background. The boundaries and scope of the thesis is laid out along with definitions of terminology. Summaries of all the Chapters are outlined.

Chapter 2

State of the Art Literature Review

Chapter abstract: *A state of the art research literature review was conducted in the domain of Test Mode Effect. This Chapter highlights critical gaps, raises research questions and outlines the aims of the research. "Test Mode effect", refers to the differential impact of the mode of presentation of assessment on students' performance. Equivalency of computer-generated scores in a CBA and corresponding pen & paper based test scores is a major issue in CBA research literature. It is argued here that investigation is required to assess the impact of various presentation formats of CBAs which may affect students' performance in a CBA environment. This state of the art review extends the boundaries of "test mode effect" to argue that "presentation format effect" of CBAs needs to be investigated in order to compare performance of students across the testing modes.*

2.1 Introduction

Initially when computer based tests were developed from pre-existing paper and pen based tests, information was usually required regarding the comparability of the test scores obtained in one mode to scores obtained in the other mode. Bunderson, Inouye and Olsen in 1989 had stated that for the field of computer-based testing, "The fundamental research question...is the equivalence of scores between a computerized version of a test and the original version". Comparability of test scores across CBA (Computer Based Assessment) and PPT (Pen and Paper based Test) is vital for establishing validity across testing modes and ensuring that norms developed from a paper and pen administered version are usable for the computer based version of the identical test as well. Moreover, it is important for providing fair testing to all examinees. According to the Joint Committee's Standards for Educational and Psychological Testing (1985), when alternate forms of an examination are in use, "it should be a matter of indifference to anyone taking the test or to anyone using the test results".

Comparability studies analyze the effects on student performance of the mode of test delivery, i.e., either computer- or paper-based mode; these are referred to as "mode effects". Though universities across the globe are switching to CBA for the unique advantages it provides over the traditional pen and paper based testing environment, still in high-stakes educational testing the impact of CBA is minimal (Underwood & Brown, 1997). The gradual nature of implementation of CBA in high-stakes educational testing can be attributed to the limited access to technology, conservatism in many education systems and resistance to some of the limitations CBA imposes (e.g. difficulties in scoring open-ended questions) (McDonald, 2002). As a result of this it is obvious and likely that CBA and P&P (Pen and Paper based) tests would co-exist for the foreseeable future, with some tests existing in both formats.The issue of equivalence is therefore very significant and, although has been argued by some to be a short-

term problem (e.g., Sutton, 1997), is probably yet to come to the fore. The main reasons for the need to establish equivalence as highlighted by McDonald (2002) are as follows:

i. In a situation, where a P&P test has been translated into CBA form or where parallel CBA and P&P tests are in use, it is necessary to establish score and construct equivalence if scores from the two are to be interchangeable. In the case of a translated test, established test statistics are required to be used. This is particularly important in high-stakes educational testing and testing for diagnostic purposes where cut-scores are established (e.g., the Joint Entrance Examinations (JEE) conducted by the Indian Institute of Technologies (IITs) were taken by 14 lakh students in 2013 – now to be placed in a CBT format (Gohain, 2012; Rao 2013)) .

ii. Research on educational standards needs to establish the equivalence between P&P and CBA if it is to report adequately on changes over time, as data may be primarily collected using CBA in the future. The rapid changes in ICT also mean that factors that are hypothesized to affect CBA performance (e.g. computer familiarity) need to be taken into account in this work.

iii. During the twentieth century a vast amount of theoretical and empirical work were conducted on the nature of human abilities in P&P tests. The results of these studies provided excellent indicators regarding the variable effects of P&P test format on human abilities. Attempts to establish construct equivalence can therefore provide valuable evidence on whether the abilities being assessed through CBA are the same as those assessed through P&P.

Previous investigations of test-mode effects have been undertaken in a variety of contexts. For instance, Federico (1989) focused on a test for recognizing fighter-plane silhouettes and Mazzeo, Druesne, Raffeld, Checketts and Muhlstein (1991) focused on the Mathematics and English College Level Examination program (CLEP). In both cases, scores on the paper-based test were superior to the computer-based test scores.

However, other studies on the test-mode effect (Mason, Patry & Berstein, 2001; Schaeffer, Reese, Steffen, McKinley, & Mills, 1993) have reported no differences between computer and paper-based tests. These kind of contradicting findings have intensified the investigations of test mode effect across the computer based platform and the traditional pen and paper based platform in order to substantiate the effect of both the testing modes so that an optimized framework for comparing the testing modes can be established. This chapter reports a state of the art review on “test mode effect studies” and presents the case of investigation undertaken in this thesis based on the research gaps identified.

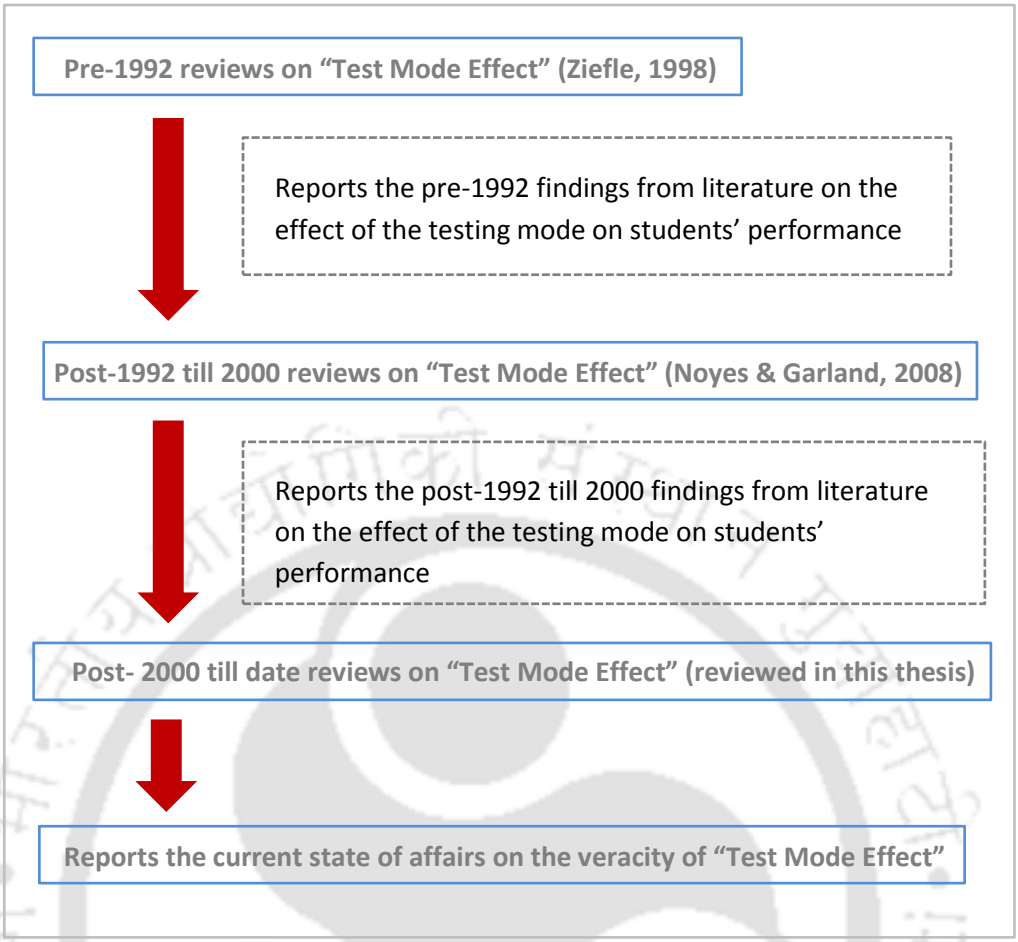


Figure 2.1 Flow diagram of the state of the art review process adopted in this thesis for reviewing the existence of test mode effect findings.

2.2 Previous studies on test mode effect

Theoretically identical paper based and computer based exams should fetch same scores but in reality there are differences as reported by Clariana R. and Wallace P. (2002) and Goldberg A. and Pedulla J. (2002). This differential impact of the mode of presentation of assessment on students' performance is referred to as *"test mode effect"* (Özalp-Yaman & Cağıltay, 2010). Many researchers have identified individual characteristics as a predictor of test mode effect. The graphical representation of the review process adopted in this thesis is depicted in the figure 2.1.

Researchers like Parshall, Spray, Kalohn and Davey (2002) suggest that equivalency and comparability of paper based tests and computer based testes should be checked before opting for a computer based testing format. Previous investigations of the test mode effect have occurred in a variety of contexts. For instance Federico (1989) carried out a test for recognizing fighter plane silhouettes and found that paper based test scores were superior to computer based test scores. Similarly Mazzeo, Druesne, Raffeld, Checketts and Muhlstein (1991) investigated

testing formats on the Mathematics and English CLEP and found that paper based test scores were superior to computer based test scores. Contrary to this when DeAngelis (2000) carried out a test for a dental hygiene course unit midterm examination, the results were opposite. Computer based test scores were higher than paper based test scores.

Other studies on the test mode effect such as by Mason, Patry and Berstein in 2001 and Schaeffer, Reese, Steffen, McKinley and Mills in 1993 have reported no difference between computer and paper-based tests. From these conflicting results, it is clear that the literature and the researchers in the area of test mode effect are divided on the issue of the equivalency of test scores from an identical paper based and computer based test modality. This is more evident from the meta-analysis conducted by Bundersonet.al (1989) where they reported the results of twenty three studies. Three studies were found to have shown the superiority of computer based test scores over paper based tests, eleven studies showed no difference, and nine studies showed superiority of paper based tests over computer based tests.

Overall, all these studies have highlighted a significant problem that test scores from identical CBA and pen and paper based assessment are not equivalent. They concluded that though paper based test scores were often higher but the difference between the paper based tests scores and the CBA scores were quiet small. Similarly, Mead and Drasgow (1993) conducted a separate meta-analysis of computer versus paper based cognitive ability tests and they concluded that on average, paper-based test scores were slightly greater than computer-based test scores.

Literature in educational technology research that might help illuminate such inconsistencies in research findings can be broadly classified into three categories:

i. Research on media (e.g., video versus text)
ii. Comparing methods (i.e., instructional strategies) and
iii. Investigations into the impact of learner characteristics (e.g., effects of ability on achievement).

Studies particularly in the area of learner characteristics have revealed that in a hypermedia environment individual learner characteristics can influence test performance. For example, Dillon and Gabbard (1998) studied the effect of learning comprehension, learner control, and style (i.e., learner characteristics) in Hypermedia and concluded that there is substantial evidence that individual characteristics affect learners' success in hypermedia environments. Extending this argument to test mode effect studies it can be stated that though paper based versus computer- mediated instructional components should produce exactly equivalent results if the contents and the cognitive activities are identical (Clarke, 1994), the discrepancy in the practical domain may be attributed to the learner characteristics that interact

with the mode of presentation to affect examinees' test performance. Therefore the amount of flexibility the test mode provides the examinee to use his / her natural behaviour while interacting with the media may have a sizeable impact on the overall success rate. So investigating individual characteristics of learners may provide ample opportunities for determining the key elements involved in the test mode effect (Wallace and Clariana, 2000; Watson, 2001).

In the study, Wallace and Clariana (2000) completed a program evaluation of a four-week long spread sheet module converted to online delivery and concluded that learners who were less familiar with the content, with computers, and who are competitive would not do well online, and so these learners should be allowed or even required to take that module in the traditional classroom setting. Similarly, Watson (2001) reported that students with higher academic attainment and also those with greater frequency of computer use benefited most from computer-aided learning.

Earlier studies (pre-1992) of comparison of computer based and paper based tests had highlighted that students in the paper based format performed significantly better according to the metrics of speed, accuracy and comprehension (Noyes & Garland, 2008). The inconsistencies in the earlier findings have been highlighted by Ziefle (1998) who attributed it to the variations in visual quality of the two formats of presentation of the test items. Ziefle (1998) highlighted that as result of the display characteristics of the computer media the eyes tire more quickly and hence this may have a pronounced effect on the performance of the students in both the testing conditions. It important to state that display characteristics of the computer media has changed dramatically over the years from early 1980s to 1990 when the previous studies had been carried out. Probably more recent studies can provide with much more compelling evidence regarding the comparability of the testing modes.

Noyes and Garland (2008) in their review has highlighted how inconsistencies are still shown in investigations in "test mode effect" studies post-1992. While studies conducted by Mason, Patry and Bernstein (2001), Noyes and Garland (2003), van De Velde and von Grünau (2003), Bodmann and Robinson (2004) and Garland and Noyes (2004) failed to find any difference in the performance of the students measured in terms of correct answers across the two testing modes while Wästlund, Reinikka, Norlander, and Archer (2005) found that comprehension from paper was superior (in terms of quantity not quality) in comparison to comprehension from the computer based mode. One of the important perspectives that were highlighted in the review was the difference in the metrics that were used as criteria to evaluate the performances of the two testing modes prior to 1992 and post 1992. It was observed that while reading speed and accuracy was used as metrics to evaluate the performance of students across the two testing modes in the studies prior to 1992, more refined metrics were used for task specific performance and evaluation in the comparability studies post 1992.

Advancement in Information & Communication Technologies (ICT), over the years has improved the computing system. Major changes were made in the display unit of the computing system as well as in the nature of the input mechanisms that were in use in traditional computing system. As a result of the changes in the computing environment, the influence of these features of the computing system cannot be ruled out without an empirical investigation that requires measures with a greater degree of sensitivity and / or studies that examine performance in the longer term. Traditional metrics employed by researchers prior to 1992 to capture the performance difference across the testing modes might therefore fail to address these issues empirically. It is evident from studies conducted post-1992, especially by Noyes and Garland (2003) and Garland and Noyes (2004), that differences exist in the manner in which the information was retrieved between the two modes of presentation. This suggested that there were differences in the memory processing dependent upon the nature of the visual input. Similarly, Wästlund et al. (2005) provided a psycho-physiological explanation regarding the differences they observed in the consumption of information (measured in terms of reading comprehension) and production of information across the testing modes. These effects were attributed to an increase in cognitive demands, that is, the computer task was more tiring and more stressful than the paper-based task and this led to 'a greater mobilization of both perceptual and executive cognitive resources' being invested (Wästlund et al. 2005). Though the study did not specifically measure cognitive load, it concluded that the VDT (Virtual Display Terminal) resulted in a higher cognitive workload.

Cognitive workload (see Appendix 4, page 240) and performance was measured and compared across testing modes by Noyes et al. (2004) in their study. They found no significant difference in the comprehension scores or the overall workload scores for the two media; significantly more workload was reported on the effort dimension for the computer-based task. Studies conducted post- 1992 had provided a greater insight into the effect of the testing mode on the students from the perspective of sensitive metrics such as cognitive workload and memory retrieval indicators that highlight how the experiences of the students differ across the testing modes. Table 2.1 (presented in the next page) encapsulates the findings of some of the studies carried out post-2000 in the area of test mode effect. It is evident from table 2.1 that contradictions still exist on equivalency of test scores across CBA and PPT. While studies conducted by Mason, Patry and Bernstein (2001), Bodmann and Robinson (2004), Noyes, Garland and Robbins (2004), Poggio, Glasnapp, Yang and Poggio (2005), Folk, March and Hurst (2006), Kim and Huynh (2007), Kim and Huynh (2008) and Tsai and Shin (2012) could hardly find any significant differences in test scores across the two modes; DeAngelis (2000), Clariana and Wallace (2002), Goldberg

Table 2.1

Test mode effect studies carried out post-2000.

Sl. No.	Studies	Subject	Findings	Test mode rated as difficult
1.	DeAngelis (2000).	Dental hygiene	CBA participants scored significantly higher than the paper group.	Paper
2.	Lee &Weerakoon (2001)	Microbiology	Students 'cohort in PBT performed significantly better than the CBT. But no significant difference between rankings across the two testing modes.	Computer in terms of test scores across the two modes but comparable when ranking was used as performance measure across modes.
3.	Mason, Patry & Bernstein (2001).	Introductory Psychology	No significant difference was found between test scores across the two testing modes.	Comparable
4.	Clariana & Wallace (2002)	Computer Fundamentals	CBT group out performed PBT group.	Paper
5.	Goldberg & Pedulla (2002)	Graduate Record Exam (GRE)	Examinees in the PBT group outperformed the CBT without-editorial-control group on all subtests.	Paper
6.	Bodmann & Robinson (2004)	Educational psychology	Completion rate of test items was faster in CBT as compared to PBT with no difference in test scores.	Comparable
7.	Lee (2004)	English essay composition in 50 minutes	No differences in holistic ratings, but analytical components marked higher in computer-generated essays.	Analytical components of Paper based essays
8.	Noyes, Garland & Robbins (2004).	Comprehension task with 10 multi-choice items	No significant difference incomprehension task, but difference in workload with the CBT required more effort.	Comparable, but CBT required more effort.
9.	Poggio, Glasnapp, Yang & Poggio (2005)	Mathematics	No meaningful statistical differences in the test scores across the two testing modes.	Comparable

Table 2.1 (contd.)

Test mode effect studies carried out post-2000.

Sl. No.	Studies	Subject	Findings	Test mode rated as difficult
10.	Wallace & Clariana (2005)	Computer Fundamentals	Performance on the CBT was significantly greater than PBT.	Paper
11.	Folk, March & Hurst (2006)	Veterinary medical education	Results indicate no significant difference in test scores across the two test modes.	Comparable
12.	Kim & Huynh (2007)	Algebra & Biology	Scores were comparable across the two testing modes.	Comparable
13.	Bennett, Braswell, Oranje, Sandene, Kaplan & Yan (2008)	Mathematics	CBT was significantly harder statistically than PBT.	Computer
14.	Kim & Huynh (2008)	English	The overall results indicated that scores obtained from PBT and CBT were comparable.	Comparable
15.	Jeong (2012)	Korean Language Mathematics Social Science Science	Significant differences between the PBT and CBT scores for the Korean language and science tests.	Computer
16.	Tsai & Shin (2012)	Dental Hygiene	Test scores were comparable across both the test modes.	Comparable

and Pedulla (2002), Lee (2004) and Jeong (2012) found students performed better in paper and pen mode than the computer based mode. Two studies that found test scores of the computer based mode outperforming the traditional pen and paper based mode was that of Wallace and Clariana (2005) and Bennett, Braswell, Oranje, Sandene, Kaplan and Yan (2008). It is important to highlight that majority of the studies post-2000 used test score and ranking metrics alone to investigate the performance difference of the students across the two testing modes. As discussed above it is imperative that in order to optimize a test mode to achieve near equivalent test format across modes, sensitive metrics should be captured so that the effect of the testing mode can be clearly elicited. Moreover, earlier studies in test mode effect seem to ignore the effect of the GUIs designed by user interface designers on the students, during the comparability studies.

Design of information systems demands focusing on users' needs and their expectations so that they can be persuaded to use the system comfortably again and again. The students' first form an initial expectation of a specific computer based assessment system prior

to its usage. They have their own mental model (see Appendix 4, page 241) regarding the workflow of the system, if the designer fails to incorporate students' expectations it can severely affect their performance and satisfaction. None of the studies discussed above have captured the extent to which the CBA system presented to them have performed in terms of their expectations. To analyse the potential effects of the presentation format of a designed User Interface it is necessary that sensitive metrics such as usability of the system, perceived satisfaction should also be measured apart from the test scores. Moreover, cognitive workload as highlighted by Noyes et al. (2004) should also be considered as an effective metric for comparative analysis. It is supported by other researchers (Vincent, Craik, & Furedy, 1996) that cognitive workload can distinguish differences in processing and that this measure is able to identify small yet important variations in performance, which is especially relevant in more sophisticated tasks that may require sustained attention, decision making, problem solving etc. In this regard the framework suggested by Paas and van Merriënboer (1993) which provides us with a computational approach to combine measures of mental effort with measures of the task performance to compare the mental efficiency of different instructions given in CBAs can be utilized. This, of course, could also be adopted for different presentation formats.

Thus according to this approach, high-task performance that is associated with low effort is called high-instructional efficiency, whereas low-task performance that is associated with high effort is called low-instructional efficiency. The studies discussed above compares the traditional pen and paper based testing mode with the computer based testing mode, while it is posited here in this thesis that in order to comprehend the effect of the CBA system more detailed way, empirical investigations need to be conducted across various presentation formats of the same testing mode. It is important to conduct such a study because CBA system in comparison to the traditional pen and paper based test provides more opportunities to the testers and the designers for innovative testing formats using various interface features. Therefore it is imperative to identify the effect of the various interface features on the students in order to establish an optimized set of features that are more suitable for such specific testing situations.

2.3 Factors affecting equivalence of test scores

There are sufficient evidences to argue that though identical test content in both the traditional PPT and the CBA should fetch the same results, it's not the case. Studies reviewed above supports the fact that there are differences in the performances of students in identical pen and paper based format and the CBA format respectively. Statistical investigations of equivalence have largely ignored the fact that in presenting a test on computer, a qualitatively different testing experience is created (McDonald, 2002). Figure 2.2 (next page) depicts McDonald's graphical representation of the equivalence of CBA and PBT. This idea of

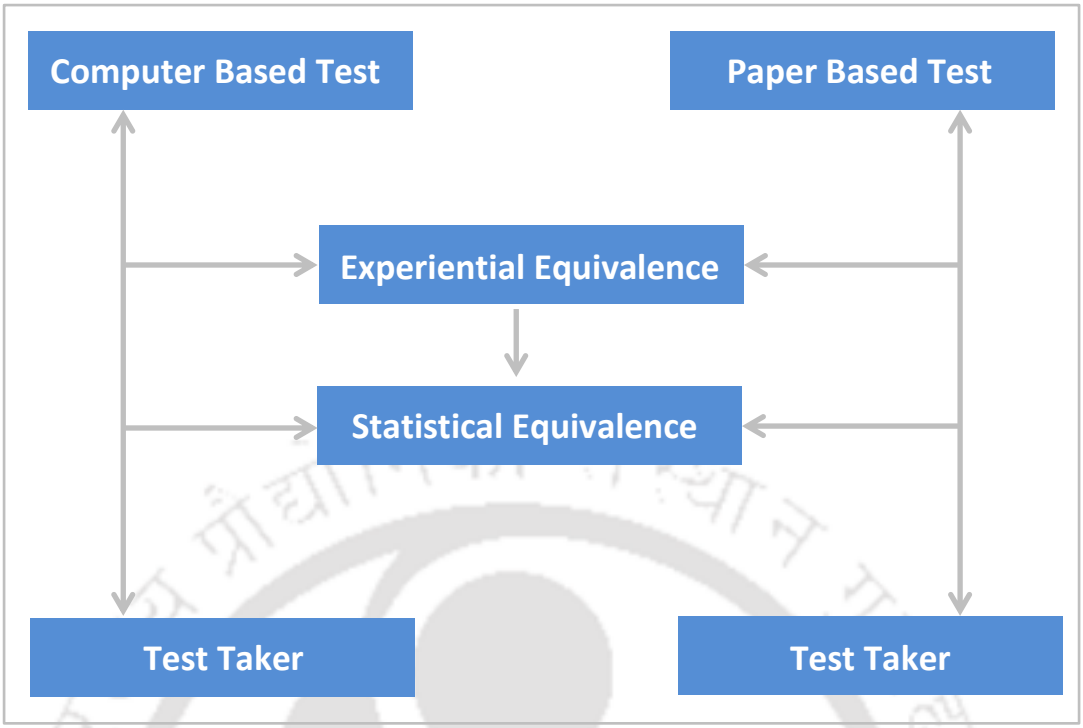


Figure 2.2 McDonald's (2002) graphical representation of statistical and experiential equivalence of Paper-and-pencil and computer-based tests.

capturing the experiential part of students in a particular assessment mode is also supported by Honaker (1988) who argued that there is a need to look at 'experiential' equivalence, making the reactions of individuals to the assessment experience a valid source of study in the field of equivalence. Though this argument was not accepted by many researchers who stated that equivalent CBA and P&P tests should not differentially affect individual differences (Tseng, Tiplady, Macleod, & Wright, 1998) however the impact of the testing modes on individual differences such as cognitive load demonstrated by Noyes et al. (2004) highlights the differential impact of the testing modes on the students. McDonald (2002) presented a model (Fig. 1) explaining how the testing experience can impact on the statistical equivalence of CBA and P&P tests.

McDonald (2002) argues that performance of students in test results from an interaction between the test taker and the test itself. The interaction between the test taker and the test gives rise to the individual's experience of the testing situation which can vary according to a range of factors including the format of the test and the students' reactions to it. If these individual reactions differ as a result of the varying test format, they will affect the constructs measured by each test and so their statistical equivalence. Reactions to tests will stem from individual differences between test takers and can take many forms and operate at different levels (e.g. individual children within a class, at the cultural background level or countries).

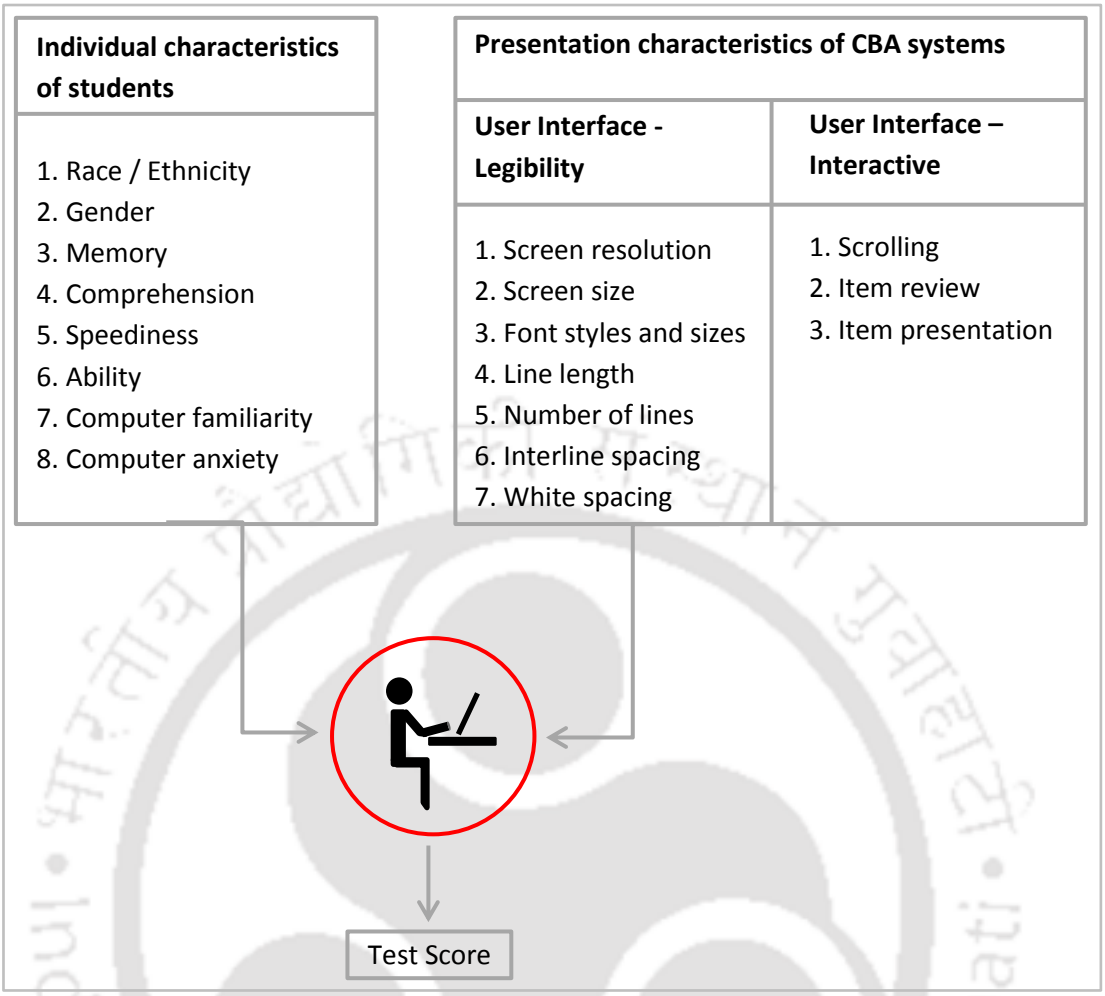


Figure 2.3 Schematic diagram of factors affecting students' performance in a test mode.

CBA and PPT has marked differences in terms of the physical dimensions. For example, paper-based assessments typically consist of pages of questions, whereas questions are often presented individually in CBA. The differences in the physical properties which results in difference in the presentation format between the testing modes may affect students' individual characteristics thereby affecting their test performance. Leeson (2006) in his review highlights these potential factors that may affect equivalency of testing modes. Fig. 2.3 presented above highlights the various individual characteristics and presentation characteristics of the CBA system that have been shown investigated in research literature on test mode effect studies by Leeson (2006). It is important to highlight here that the objective of the current research work is to empirically investigate the effects of various presentation formats on students' performance. User Interface designers depending on the characteristics of the mental model visualized by them can utilize different interface features for a desired functionality. In this thesis, therefore the focus of investigation centers around the impact of such interface features on the students in the CBA testing mode.

Lee's (2006) review of 'test mode effect studies' highlights various interface features where contradictory findings have been reported and therefore their impact on the students' remains still unanswered. While there is a converging agreement among the research community regarding the impact of the factors affecting the legibility issues in a user interface of a CBA system, the same cannot be said for the factors affecting the interactive issues in a user interface of the CBA system.

In the subsequent section a detailed review on the interactive features like,

i. Scrolling,
ii. Navigation
iii. Item review and presentation and
iv. White space

have been presented where contradictions in literature exist.

2.3.1 Scrolling

Previous research on the effect of scrolling on performance in summative CBA environments can be classified into two broad categories: (i) Studies that investigated the impact of scrolling on reading comprehension and (ii) studies that investigated the impact of scrolling on students' performance in multiple choices tests. There is a paucity of research on the impact of scrolling on students' performance in multiple choice tests and the results of existing research points in different directions.

Ricketts and Wilks (2002), for example, documented that students' performance was worse when the test questions in a CBA were presented by an interface that required scrolling in comparison to a traditional paper-pencil-format without scrolling as well as in comparison to an interface in which students received one question at a time. However, since crucial methodological aspects like sample size, significance levels and potential influential factors like students' previous experience with CBAs or students attitudes towards computers were not reported in the study, the results have to be taken into account cautiously (Charlton & Hewson (2007)). Other studies on the effect of scrolling on students' performance in CBAs with multiple choice tests found no differences in students' performance between interfaces with scrolling versus traditional paper-pencil-based format (Cassady & Gridley, 2005; Charlton & Hewson (2007), Hewson, 2012). It is important to highlight here that students were given ample time to complete the tests in the study conducted by Hewson et al. (2007) while rest of the studies offered students with limited time settings. This might have diluted the effect of interface features on students' performance. In a time limited setting format the effect of the interface features may be pronounced. Not only is the existing scarce research on the effect of scrolling in computer administered multiple choice tests characterized by methodological problems, but also

we in this thesis think that a different empirical approach would be beneficial to analyze the effect. In the subsequent section we will outline why we think this is the case.

Previous studies (Charlton & Hewson (2007); Whitelock, 2009) mostly focused the effect of scrolling and item presentations on students' performance tests. These studies did not find a significant difference in performance (test scores) between groups with and without scrolling (Charlton & Hewson (2007); Whitelock, 2009). Even though performance of the students in these tests might not be affected, scrolling and item presentation might have an effect on other variables like students' cognitive load (Noyes et al., (2004), which is generally affected by differences in the designs of CBAs (Brünken, Plass, & Leutner, 2003). Therefore, it is required that empirical investigations are carried out in limited time setting format to validate the effect of scrolling on students performance as well as individual characteristics of students in CBA systems.

Though scrolling is considered as feature of the navigational mechanisms but CBA systems that feature scrolling manipulate item presentation as well in comparison to other navigational mechanisms which does not affect item presentation. Item presentation is an important user interface feature that is addressed by scrolling. While all the items are presented simultaneously in a CBA system that features scrolling, other navigational mechanisms restricts the presentation of the test items to one item per screen. Since, a considerable amount of studies in literature of test mode effect have been devoted on scrolling, therefore a separate review on scrolling has been presented here. Though a review on navigational mechanisms in test mode effect studies have been presented in the subsequent section but it is devoid of the reviews of scrolling which behaves more than a navigational mechanism due to its unique characteristics of addressing item presentations in CBA systems.

2.3.2 Navigation

Unlike hypermedia based learning environments, summative computer based assessment environments which are a component of online learning environments are high stress situations that requires exhaustive use of cognitive resources in limited time setting formats by students. Therefore, it is important that the computer based delivery media of the assessment content does not interfere with the problem solving activity of the students. The computer based delivery media must act as a facilitating tool supporting the students in focusing their cognitive resources in problem solving activity. If the delivery mode forces the students to spend time and cognitive resources in learning the tool itself then performance of the students may be compromised. Literature highlight studies that investigated the effect of the presentation mode such as screen size and resolution, scrolling, item review and item presentation on the

performance of the students in a computer based assessment environment (Leeson, 2006). There have been rarely any studies that investigated the effect of navigational mechanism in CBA environment. Moreover, the study of navigational mechanisms in a computer based assessment environment assumes more significance because if the activities of the students like their ability to plan their actions, determining their location, and find their way back if they follow an inappropriate navigation item, are compromised; it may lead to disorientation and increase in cognitive overload affecting their performance especially in stressful examination context.

Effects of navigational mechanisms on learners have been a major area of study in hypermedia based learning environments. It has been reported that navigational tools increase navigation efficiency, reduce the feeling of being lost and improve learning performance (Allison & Hammond, 1989; Dee-Lucas & Larkin, 1995; McDonald & Stevenson, 1998; Puntambekar et al., 2003). But it may also impede the cognitive flexibility of the learners by imposing a simplified structure over the existing knowledge structure (Gall & Hannafin, 1994) and as such may not compel the learners to fully interact with the hypertext environment (Jonassen, 1986). Studies conducted by Dee-Lucas and Larkin (1995) and McDonald and Stevenson (1998) revealed the benefits of providing content list on learners navigation and memory in hypertext environments. Likewise, concept maps were found to improve learner control and cognitive flexibility in a hypertext learning environment (Novak & Gowin, 1984; Jonassen, 1986; Puntambekar et al., 2003). Apart from studies in hypermedia based learning environments there are studies that report the effect of navigational mechanisms on users' performance and subjective preference in information systems. Leuthold et al. (2010) reports the positive effects of vertical menus, that displays all the links at the same time over dynamic menus that represents condensed display and only displays all the links on request by the user by mouse roll over action on users' performance and subjective preference in web pages. A similar study conducted by Fowler and Stanwick (2004) favored vertical menus in information systems over dynamic menus. Apart from scrolling not a single study has reported the effects of other navigational mechanisms that display all the links at a particular time versus condensed display of links at a particular time in the context of a CBA environment, which is a component of the online learning environment. Effects of different format of navigational mechanisms in a CBA environment demands immediate attention because if a particular type of navigational mechanism impedes the cognitive flexibility of the students, it may affect their performance and in turn affect students' test performance. This may raise questions on the validity of the tests performance reported from assessments delivered through the computer based media.

2.3.3 Item presentation and review

A paucity of research exists in regard to the effect of item presentation on performance of students in CBA environments – studies in item presentation investigated the number of items that were modulated in the presentation screen. These studies were carried out to optimize the optimal number of items that should be presented per page. Scrolling was used when the number of items increased in order to accommodate the items per page. Leeson (2006) in his review highlights the findings in the literature wherein it is stated that presenting one item per screen tends to increase errors and generated more hurried responses; however, the ability to review items may counter this detrimental effect. Multiple items on screen may have a facilitating effect allowing examinees to skip, scan, and build off previous item information. However the findings of the existing few studies (Greaud & Green, 1986; Lee, 1986) should be interpreted cautiously as the studies used outdated apparatus and technologies making conclusions on the optimal item presentation format difficult (Leeson, 2006).Thus, the effect of item presentation on students' performance in a CBA environment needs to be revisited again.

Studies that investigated the effects of item review on students' performance found that though performance remained largely unaffected by the interface feature "review option", students preferred to have the option to review available (Leeson, 2006). It is important to highlight that previous studies that investigated the effect of flexibility offered by CBA such as item review, measured performance metrics of students based on their test scores but no study investigated the effect of this feature on the cognitive processing of the students. Moreover it was hardly investigated how these interface features affect the individual characteristics such as usability and satisfaction of the students in an assessment environment. Therefore, the case of item review also needs to be relooked from the perspective of other sensitive metrics that might capture the difference.

2.3.4 White space

White space in a CBA format denotes the amount of screen visible given the width of the text passages, namely white space (Lynch & Horton, 2002). It was believed that for traditional printed text the use of white space adds not only to the attractiveness of the text, but aids in directing the viewer's attention to the regions where important information is present (Mullet & Sano, 1995). White space helps in preventing the influence of distracting, unimportant information by spatially organizing information on screen (Lynch & Horton, 2002).

Early research conducted by de Bruijn, de Mul, and van Oostendorp (1992) and van Van Nes (1986) on the effect of white space on computer displays endorses the layout logic from the printed media. Thereafter, it was important to investigate the amount of optimal white

space that should be used in computer displays. Spool, Scanlon, Schroeder, Snyder, and DeAngelo (1999) conducted usability studies to address this issue but the study had methodological and measurement issues (Leeson, 2006) for which the conclusions derived by them have to be interpreted cautiously. Though the study found that the more white space a site had, the poorer participants performed in terms of the success in finding information but the use of different display formats and inconsistencies in information presented to the participants during the experiments across the display formats made the conclusion of the study refutable.

Later on, Bernard, Chaparro and Thomasson (2000) conducted a study to empirically investigate the effect of white space on participants. The participants were asked to find errors and hyperlinks from three identical Web sites that varied only in the amount of measurable white space displayed. The study presented the participants with three conditions of white space – low, medium and high white space condition. The Low condition presented little white space, with one character space (3 mm) separating one column of text from another, and with no additional space between paragraphs. The Medium condition had four character spaces (9 mm) between each column and a blank line (9 mm) between each paragraph. The third condition, High, had eleven character spaces (19 mm) between each column, with four blank lines (19mm) between each paragraph. The study highlighted that the participants performed equally in terms of time required to find errors and hyperlinks, they were significantly more satisfied with the amount of white space found in the medium condition than the low and high white space condition presented to them.

A similar study that addressed the issue of white space was conducted by McMullin, Varnhagen, Heng and Apedoe (2002) who investigated the effect of line length and white space on the comprehension of text presented on the Web. The study highlighted the main effect of white space wherein participants obtained higher comprehension scores in the one-column presentations across both line lengths (narrow or wide). Though the study highlighted interesting findings but the study had experimental bias. The effects of line length and visual fixations might have confounded this result (Leeson, 2006).

It is evident from the findings of the past studies conducted to investigate the effect of white space within computer displays that it might have an effect on the participants. The studies conducted to investigate this issue are scarce and many of the studies had methodological and experimental issues which made their findings inconclusive. Moreover the studies that addressed this issue, studied the effect of white space in comprehension based test item layout but rarely anyone has enquired about the effect of white space in MCQ (Multiple Choice Question) format and how graphical elements embedded in the GUI to aid the students in a test format reducing the white space may affect the students.

2.4 Insight from state of the art review

The state of the art review conducted in this chapter highlights important gaps that demands further investigation to empirically validate and substantiate the effects of the presentation characteristics of the display format in a CBA system. The most significant question arises from the fact that there are inconsistencies in the findings reported regarding the effect of interface features on students. A graphical representation of findings from the atate of the art review and the methodology adopted has been depicted in figure 2.4 (a) and 2.4 (b) (next page).

When there are inconclusive evidences regarding the effect of interface features between the same display units i.e., the CBA format, it is obvious that differences would exist when CBA display format is compared with the traditional pen and paper based format. Therefore, it is first and foremost required that empirically validated investigations are conducted to identify the optimal display format in a CBA system and then probably comparing it with the traditional format would provide greater insight into the effect of the testing modes. It is therefore, posited here that investigations should be carried out to observe whether differences among the presentation format in the same mode exists or not.

Secondly, the interface features where contradictions still exist regarding their effect on students in a CBA format namely scrolling, item review and presentation as well as white space which needs to be re-visited. The inconsistencies in the investigation procedures adopted by past researchers to demonstrate the effect of these features makes the conclusions of those studies disputable and incomplete and hence the question pertaining to how they affect students in a CBA format remains unanswered. Navigation which is a critical tool in hypermedia based learning environments has rarely been investigated for their effects on students in a CBA environment as well. These anomalies in reported studies highlight the need for an empirical investigation to elicit the effects of scrolling, item review and presentation, white space as well as navigation on students in CBA format.

Apart from the gaps highlighted above, the research literature also highlight the fact that sensitive variables have been ignored while capturing the effects of the testing modes in test mode effect studies. Cognitive load which plays a major role in determining the success of instructional intervention and has been shown to degrade performance either by under load or overload had been rarely studied in case of the interface features like scrolling, item review, white space and navigation in CBA format. Noyes, et. al., (2004) reports that cognitive workload may comprise another test mode effect, which needs to be taken into consideration when administering computer-based assessments. Capturing cognitive load is important because

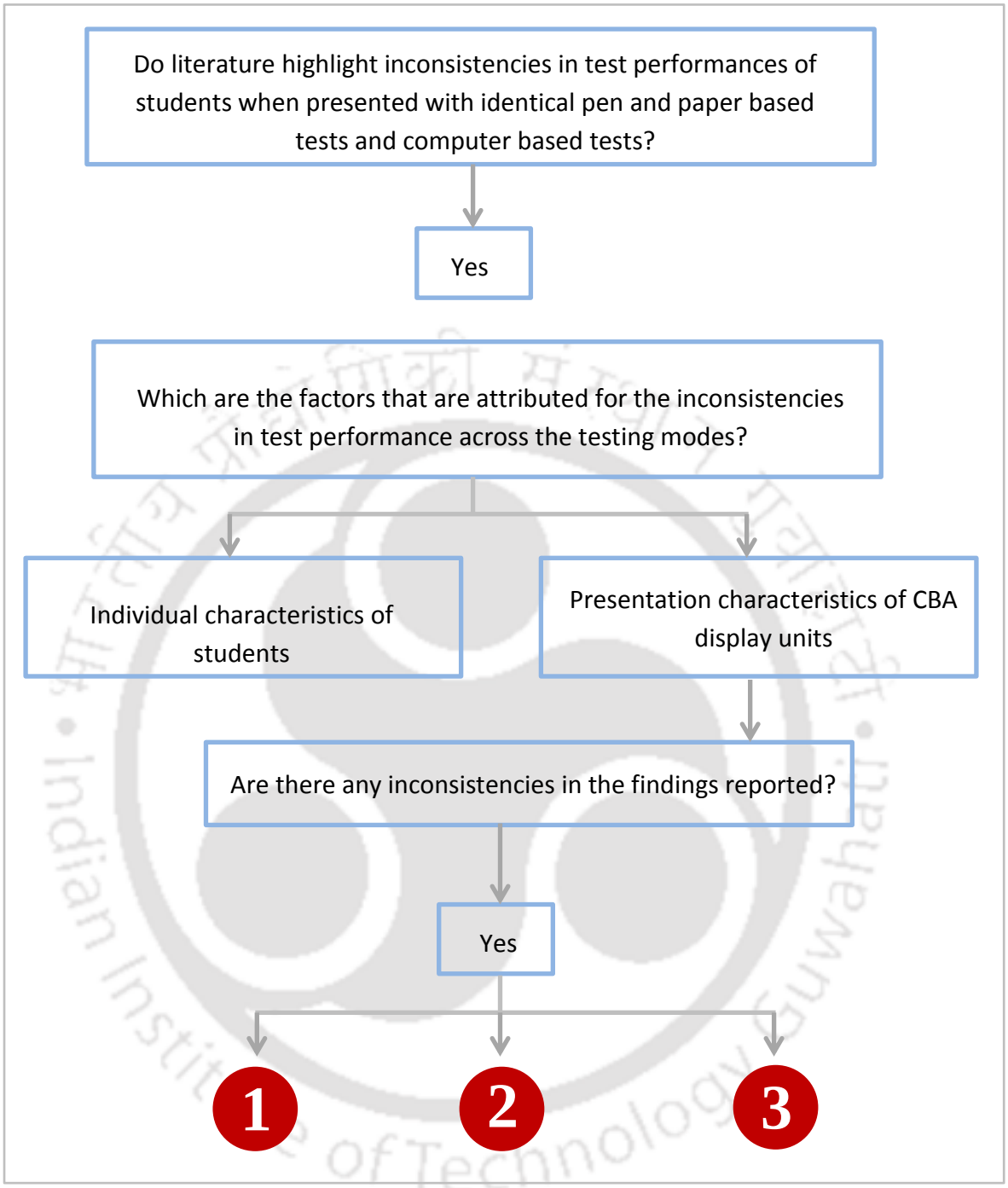


Figure 2.4 (a) Graphical representation of the state of the art review.

1

Inconsistencies in past studies on effect of CBA display units

1. *Scrolling:* Methodological issues in existing studies + Contradicting results + Ignoring the effect on sensitive variables.

2. *Item presentation & review:* Limited studies conducted + Outdated apparatus used + Ignored the effect on sensitive variables.

3. *White space:* Limited studies conducted + Methodological issues and experimental bias in existing studies + Ignored the effect on sensitive variables.

4. *Navigational mechanisms:* No published evidence found that investigated the effect of navigational mechanisms.

2

Effect of Cultural backgrounds on performance in a CBA format

No published evidence found that investigated the effect of cultural backgrounds. By cultural background we mean the influence of local contextual difference that are brought about due to difference in living styles, practice, habits etc. In this thesis students' of USA in this sense have a different cultural background as compared to the students' of India.

3

Ignoring Sensitive variables

1. Cognitive Load
2. System Usability
3. Perceived Ease of Use
4. Perceived Usefulness
5. Perceived System Quality
6. Relative Advantage
7. Perceived E-learner's Satisfaction
8. Willingness to Continue to Use

Figure 2.4 (b) Graphical representation of the findings from the state of the art review.

it can be argued that, by investing more mental effort (mental effort- the aspect of cognitive load that refers to the cognitive capacity that is actually allocated to accommodate the demands imposed by the task (Pass, et. al., 2003)) students of a particular study group might reach equal or higher performances than the corresponding study group. A particular kind of information presentation may demand more cognitive processing than another. If performance of students with both the presentation formats is equivalent, it can be argued that the presentation format that induces less cognitive load is more efficient and effective for students (efficiency view) (Ahern and Beatty, 1979). Paas and van Merriënboer (1993) suggested a computational approach to combine measures of mental effort with measures of the associated primary task performance to compare the mental efficiency of instructional conditions. They calculated the Instruction efficiency using the following formula:

$$E = \frac{M - P}{\sqrt{2}}$$

Where, E = Instructional Efficiency, M = z value of the cognitive load measure and P = z value of performance measures (test score).

They argued that the complex relation between mental effort and performance can be used to compare the mental efficiency of instructional conditions in such a way that learners' behaviour in a particular instructional condition is considered more efficient if their performance is higher than might be expected on the basis of their invested mental effort or equivalent if their invested mental effort is lower than might be expected on the basis of their performance. In this approach, high-task performance associated with low effort is called high-instructional efficiency, whereas low-task performance with high effort is called low-instructional efficiency. In this research work instructional efficiency of the CBA conditions have been calculated based on the computational formula prescribed by Paas and van Merriënboer (1993). Therefore, cognitive workload is a significant parameter to investigate the effect of interface features on performance as it can distinguish differences in processing and can identify small yet important variations in performance, which is especially relevant in more sophisticated tasks that may require sustained attention, decision making, problem solving, etc. (Noyes, & Garland, 2008).

The presentation format of a CBA system can force the students to focus their cognitive resources to learn the system itself or assist them to complete the task (problem solving activity). In a CBA environment where mastering difficult interface is a form of problem solving, some individuals adapt and learn quickly while others take more time to build their mental model of the interface (Drommi, A., et. al., 2001). Therefore, satisfaction of

Table 2.2

Dependent measures proposed for further investigation in this thesis.

Sl. No.	Dependent Measures	Definition
1.	Perceived Ease of Use	It refers to "the degree to which a person believes that using a particular system would be free of effort" (Davis, F. D., 1989).
2.	Perceived Usefulness	It is defined as "the degree to which a person believes that using a particular system would enhance his or her job performance" (Davis, F. D., 1989).
3.	System Quality	It is defined as "whether or not there are bugs in the system, the consistency of the user interface, ease of use, quality of documentation, and sometimes, quality and maintainability of the program code" (Seddon, P., 1997).
4.	Relative Advantage	It is defined as "the degree to which a new system is perceived as being better than the alternative it supersedes" (Rogers, E.M., 2002).
5.	E-learning Satisfaction	Satisfaction is an individual's feelings of pleasure or disappointment resulting from comparing a product's perceived performance (or outcome) in relation to his or her expectations (Chiu, C. M., et. al., 2005).
6.	Willingness to continue to Use	The Extent to which the student intends to continue to use an e-learning system.

students using the CBA system is an important indicator that determines the success of information systems (Delon & Mclean, 1992) in efficiently assisting the students to spend time and cognitive resources to complete the task rather than forcing them to focus on learning the tool (in this case the interface features) to complete the examination task. Hence, the interface should be highly effective so that it imposes as minimal a cognitive load as possible on the learner (Stoney & Wild, 1998). Usability tests are generally used to measure the effectiveness of interfaces (Cheon, J., & Grant, M. M., 2009). Therefore by assessing usability, its perceived quality and satisfaction of the presentation formats and associating these measures with perceived cognitive workload one can assess which format of presentation is efficient and effective in a CBA environment. It is posited here that in order to investigate the effect of interface features it is required to capture performance of study groups, their perceived cognitive load and measure usability of the systems, its perceived quality & satisfaction, so that conclusion can be drawn whether interface features are affecting students in a CBA environment or not. Apart from capturing perceived cognitive load and the usability measure of the system, the dependent measures listed in table 2.2 were proposed to be captured for assessing the impact of the interface features on perceived quality & satisfaction of the systems.

Finally, last but not the least it is required to investigate whether cultural differences among student groups can be attributed as factor for test mode. Though a number of studies

have been reported in the literature that have been carried out across various student communities but none of them carried out a cross-cultural comparison to determine the effect of the presentation format across cultural shores in a CBA format. Based on the gaps in the research literature – research questions were raised which are stated below. A detailed argument on the theoretical implications of culture as a variable that might interfere with users' (here students) responses have been elaborated in Appendix 4, page 238.

2.4.1 Research Questions

The research gaps highlighted above from the perspective of the interface features and the factors identified to capture their effects have been modelled in terms of research questions that would guide this research investigation. The research questions are listed below:

i. What is the effect of the interface features namely scrolling, navigation, item review and presentation and white space in a CBA format on the test scores of students?
ii. Do these features affect sensitive variables such as subjective cognitive load, system usability, perceived ease of use, perceived usefulness, system quality, e-learner satisfaction, willingness to continue the usage and relative advantage differentially in a CBA format?
iii. Do these interface features affect the instructional efficiency of the CBA format presented to the students?
iv. Does cultural context and backgrounds of students affect the sensitive variables such as cognitive load, system usability, perceived ease of use, perceived usefulness, system quality, e-learner satisfaction, willingness to continue the usage and relative advantage of the current CBA presentation format as compared to other CBA presentation formats they experienced earlier?
v. Is there any interaction effect of the interface feature and cultural backgrounds on the students in a CBA format?

2.5 Chapter 2: Summary

A State of the art review on literature in 'test mode effect' studies has been presented. It has been reported here that there are scientific evidences to support that test scores of identical pen and paper based tests and computer based tests are non-comparable. The factors that have been held responsible by past researchers for this inconsistency have been highlighted. The review highlights the inconclusive results of the past investigations and reports the research gaps that are required to be investigated have been outlined. The broad research questions have been stated.

Chapter 3

Research Methodology: Experiment, Hypothesis and Sample Design

Chapter Abstract: *This chapter presents the working hypotheses which have been classified in terms of the independent factors highlighted in chapter 2. A detailed description of the research methodology adopted to plan and conduct the study, has been described. The instruments for conducting the study have been described in detail. The experimental procedures of data collection both in India and in the United states have been discussed. Finally, it describes the rationale behind the statistical tools employed to analyse the data.*

3.1 Introduction

The state of the art review has highlighted that test mode affects performance of the students. It has been found that literature highlights various factors that have been investigated for affecting students' performance in the testing modes. Among these factors, interface features such as scrolling, item review and presentation, white space and navigation - require further investigation as highlighted in chapter 2. It is evident from past investigations reported in literature that sensitive variables like cognitive load, system usability, perceived ease of use, perceived usefulness, system quality, perceived e-learner satisfaction, willingness to continue the usage, relative advantage - have been ignored while measuring the performance of students in test mode effect studies.

The study reported here investigates the effect of the interface features - (i) scrolling, (ii) item review & presentation, (iii) white space and (iv) navigation - not only from the perspective of test scores and time taken to complete the test but also from the perspective of the sensitive variables which might provide us with a broader and deeper understanding about the impact of these interface features on students in a CBA (Computer Based Assessment) system.

A within mode study (CBA mode alone) has been conceived here to identify how these interface features (i to iv – stated above) might affect students' performance and their subjective experiences of the assessment mode. Such a study is important in order to identify the optimal presentation format that would enhance the students' performance and thereby helping them to achieve the learning objective. To start with the research questions, aims and objectives of the study presented in the thesis are listed below so as to give the background for unfolding the design of experiments.

3.2 Research Questions

i. Do interface features like scrolling, navigation, item review and presentation and white space in a CBA format affect test scores of students?

ii. Do these features affect sensitive variables such as – (a) perceived cognitive load; (b) perceived system usability; (c) perceived ease of use; (d) perceived usefulness; (e) perceived system quality; (f) e-learner satisfaction; (g) willingness to continue the usage; (h) relative advantage of the current CBA presentation format as compared to other CBA presentation formats they experienced earlier – in a CBA format?

iii. Do these interface features mentioned in point (i) above affect the 'instructional efficiency' of the CBA format presented to the students?

iv. Do cultural backgrounds of students affect the sensitive variables such as (a) perceived cognitive load; (b) perceived system usability; (c) perceived ease of use; (d) perceived usefulness; (e) perceived system quality; (f) e-learner satisfaction; (g) willingness to continue the usage and (h) relative advantage of the current CBA presentation format as compared to other CBA presentation formats they experienced earlier?

v. Do the interface feature and cultural backgrounds interact with each other and affect students' performance and their subjective evaluation of the assessment system in a CBA format?

3.3 Aim of the study

To determine whether the effect of Graphical User interface features and cultural backgrounds of students in a summative computer based assessment can be attributed as factors that affect students' performance and their subjective evaluation of the assessment system in a CBA mode.

3.4 Objectives of the study

i. To establish the effect of delivery media (Graphical User Interface) and cultural background of students in an e-assessment environment on student's performance.

Table 3.1

Experiment details of the studies conducted.

Experiments	Interface features
Experiment 1: Scrolling	Scrolling versus without scrolling
Experiment 2: Navigation	Vertical menu based versus step based
Experiment 3: Item review and presentation	Question customization versus Time customization
Experiment 4: Item review and presentation	Single question without customization versus Question customization
Experiment 5: Screen Density (White space)	Low screen density (Maximum white space) versus High screen density (Minimum white space)

ii. To establish the effect of delivery media (Graphical User Interface) and cultural background of students in the success of e-assessments.

iii. To determine the effect of delivery media (Graphical User Interface) and cultural background of students in an e-assessment environment on perceived subjective evaluation of the CBA system.

3.5. Research Plan: Materials & Methods

The overall experiment design adopted in the thesis is shown in figure 3.1 (next page). We posit that due to the effect of delivery media even performance of 'Experts' will be affected. Expert Users and Novice Users are defined by Prümper, Frese, Zapf and Brodbeck (1991) as "All in all, there are usually two (strongly overlapping) criteria used for the differentiation between novices and experts: knowledge (e .g. comparing students and teachers) and the time spent working with a particular system (e .g. students with a few vs. those with many courses)". Here, we have classified students on the basis of their computer attitude and their computer anxiety. Those students' who report higher computer attitude, lower computer anxiety and had previous experiences with CBA systems (more than once) and had undertaken courses or exams previously as a part of their curriculum program, have been classified as expert users. Five experiments were conducted based on the interface features (i) scrolling; (ii) navigation; (iii) item review and presentation (iv) screen density (white space). In each of the five experiments interface features had two levels. The details of the experiments are provided in table 3.1 presented above.

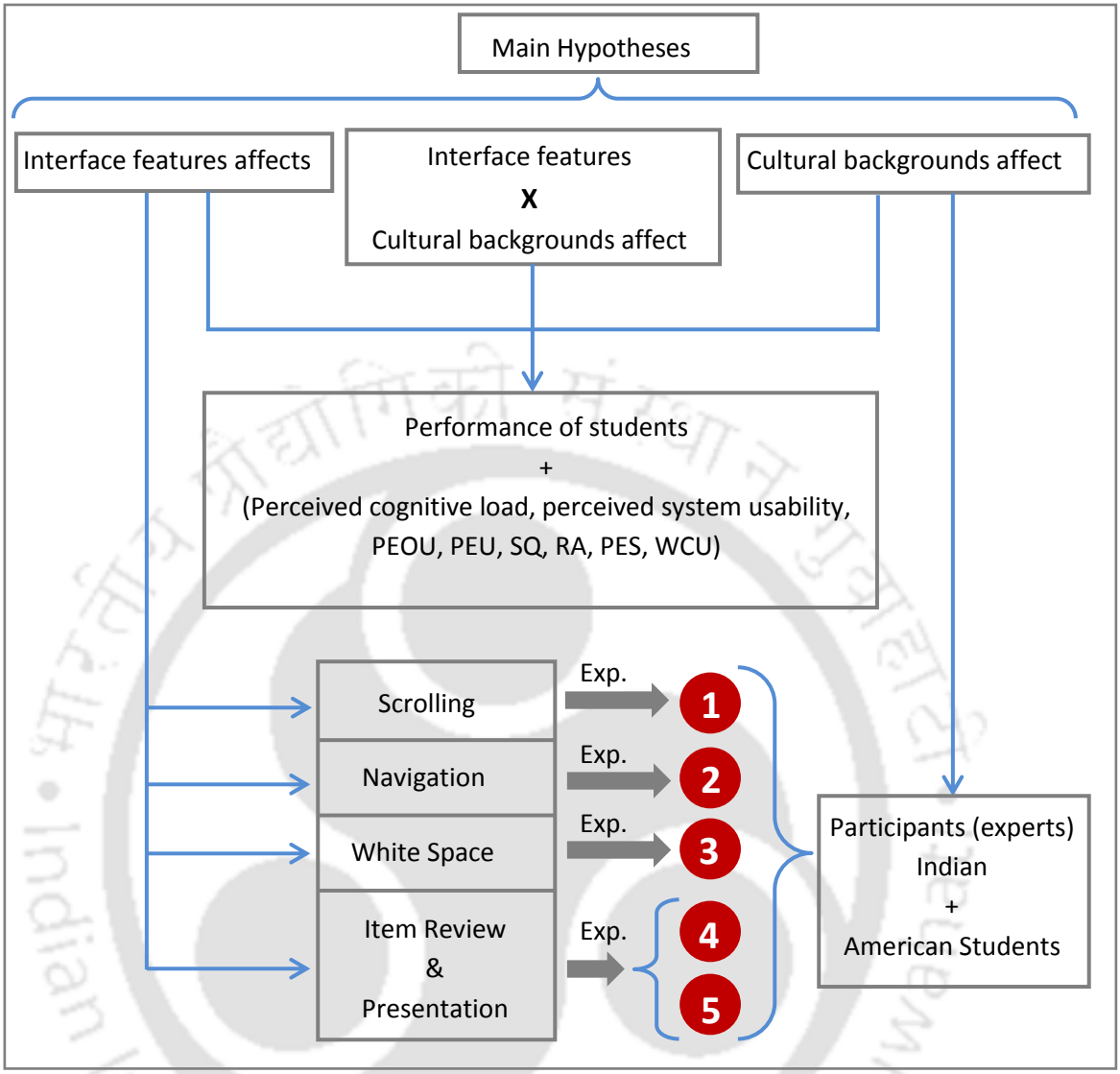


Figure 3.1. Graphical representation of the experiment design.

The posits that were investigated are listed below.

3.5.1 Posits under test in the study

The purpose of this study was to examine the effect of user interface presentation and cultural backgrounds of students' on their performance, self-reported cognitive load, perceived usability, perceived ease of use, perceived usefulness, perceived system quality, relative advantage, e-learning satisfaction and students' willingness to continue to use the system. The following posits were formulated and forms the basis for the working hypotheses discussed in the next section.

P_1: *Interface features* alone would affect performance of students, perceived cognitive load, perceived system usability, perceived ease of use, perceived usefulness, system quality, e-learner's

satisfaction, willingness to continue the usage and relative advantage in a summative CBA environment.

P_2: *Cultural background* alone would affect performance of students, perceived cognitive load, perceived system usability, perceived ease of use, perceived usefulness, system quality, e-learner's satisfaction, willingness to continue the usage and relative advantage in a summative CBA environment.

P_3: The interacting effect of *cultural background* and *interface feature* alone would affect performance of students, perceived cognitive load, perceived system usability, perceived ease of use, perceived usefulness, system quality, e-learner's satisfaction, willingness to continue the usage and relative advantage in a summative CBA environment.

3.5.2Working hypotheses

To investigate the effects of interface features and cultural backgrounds, five experiments were conducted. Each of the five experiments investigated the effect of a particular interface feature and the effect of cultural backgrounds on the students' performance and their subjective evaluation of the CBA system. The criterion that influenced the choice of the particular features for navigation and item review and presentation experiments is listed below:

(i) Navigation mechanisms: Vertical menu based navigation and step based navigation are the two different navigational mechanisms in use in the two most widely taken computer based online test in India, namely BITSAT (Birla Institute of Technology and Sciences Aptitude Test) conducted by Birla Institute of technology and CAT (Common Admission Test) conducted by the IIMs (Indian Institute of Management). While BITSAT implemented step as well as vertical menu based navigation mechanism, CAT implemented step navigation only in their computer based online aptitude test system. The current study is based on the features that are in use in these online computer based tests.

(ii) Item review and presentation: Question customization and Time customization - Dhar and Yammiyavar (2011) highlighted in a pen and paper based assessment, majority of the Indian students prefer to see all the test items the moment the assessment period starts and then they decide the order of answering questions. Moreover, it was also highlighted by Dhar and Yammiyavar (2011) that time is an important constituent in the assessment period. Often, students decide the amount of time needed to be invested in a particular question depending on the complexity of the test item. The results and the conclusion of the study reported by Dhar and Yammiyavar (2011)

influenced the design of the interface features in the experiment 3 and 4. In experiment 3 students were allowed to view all the questions with the answer options, but were not allowed to answer. They were allowed to review the test items for few minutes and decide order of the questions to be presented during the main assessment period. Since, the student could customize his/her order of test items during the main assessment period the experiment was named question customization.

In the time customization experiment students were allowed to view all the test items simultaneously. The assessment session comprised of two sections. The students could view test items of both the sections simultaneously. Later they were asked to allot time to both the sections out of the total time allotted for the assessment session. Hence, the name 'Time customization' was coined for the experiment.

In the item review and presentation literature, it was highlighted that studies focussed on reviewing options primarily from the perspective that once the students answered the test items they were allowed to review their answers before final submission. A number of studies in item presentation investigated the optimal number of items that the students prefer to see during a CBA session. The current investigation takes a different approach in item review and presentation. Here, students were not only allowed to review their answers but were also allowed to customize test item's order for the main assessment session or customize time for the all the sections for the main assessment session. Based on these five different experiments that were planned, working hypotheses were formulated for each one of them. The working hypotheses are presented in the table 3.2 (next page).

Table 3.2

Hypotheses being tested under each of the five Experiments.

Experiment	Working Hypotheses
1. Scrolling versus without scrolling	**H_{1a}**: Scrolling alone would affect performance of students, perceived cognitive load, perceived system usability, perceived ease of use, perceived usefulness, system quality, e-learner's satisfaction, willingness to continue the usage and relative advantage in a summative CBA environment. **H_{1b}**: Cultural background alone would affect performance of students, perceived cognitive load, perceived system usability, perceived ease of use, perceived usefulness, system quality, e-learner's satisfaction, willingness to continue the usage and relative advantage in a summative CBA environment. **H_{1c}**: The combined interacting effect of cultural background and scrolling alone would affect performance of students, perceived cognitive load, perceived system usability, perceived ease of use, perceived usefulness, system quality, e-learner's satisfaction, willingness to continue the usage and relative advantage in a summative CBA environment.
2. Vertical menu based navigation versus step based navigation	**H_{2a}**: Navigation alone would affect performance of students, perceived cognitive load, perceived system usability, perceived ease of use, perceived usefulness, system quality, e-learner's satisfaction, willingness to continue the usage and relative advantage in a summative CBA environment. **H_{2b}**: Cultural background alone would affect performance of students, perceived cognitive load, perceived system usability, perceived ease of use, perceived usefulness, system quality, e-learner's satisfaction, willingness to continue the usage and relative advantage in a summative CBA environment. **H_{3c}**: The combined interacting effect of cultural background and navigation alone would affect performance of students, perceived cognitive load, perceived system usability, perceived ease of use, perceived usefulness, system quality, e-learner's satisfaction, willingness to continue the usage and relative advantage in a summative CBA environment.
3. Question customization versus Time customization	**H_{3a}**: Customization feature alone would affect performance of students, perceived cognitive load, perceived system usability, perceived ease of use, perceived usefulness, system quality, e-learner's satisfaction, willingness to continue the usage and relative advantage in a summative CBA environment. **H_{3b}**: Cultural background alone would affect performance of students, perceived cognitive load, perceived system usability, perceived ease of use, perceived usefulness, system quality, e-learner's satisfaction, willingness to continue the usage and relative advantage in a summative CBA environment. **H_{3c}**: The combined interacting effect of cultural background and customization feature alone would affect performance of students, perceived cognitive load, perceived system usability, perceived ease of use, perceived usefulness, system quality, e-learner's satisfaction, willingness to continue the usage and relative advantage in a summative CBA environment.

Table 3.2 (contd.)

Hypotheses being tested under each of the five Experiments.

Experiment	Working Hypotheses
4. Single question without customization versus Question customization	**H_{4a}**: Customization alone would affect performance of students, perceived cognitive load, perceived system usability, perceived ease of use, perceived usefulness, system quality, e-learner's satisfaction, willingness to continue the usage and relative advantage in a summative CBA environment. **H_{4b}**: Cultural background alone would affect performance of students, perceived cognitive load, perceived system usability, perceived ease of use, perceived usefulness, system quality, e-learner's satisfaction, willingness to continue the usage and relative advantage in a summative CBA environment. **H_{4c}**: The combined interacting effect of cultural background and customization alone would affect performance of students, perceived cognitive load, perceived system usability, perceived ease of use, perceived usefulness, system quality, e-learner's satisfaction, willingness to continue the usage and relative advantage in a summative CBA environment.
5. Low screen density (maximum white space) versus High screen density (minimum white space)	**H_{5a}**: White space or screen density alone would affect performance of students, perceived cognitive load, perceived system usability, perceived ease of use, perceived usefulness, system quality, e-learner's satisfaction, willingness to continue the usage and relative advantage in a summative CBA environment. **H_{5b}**: Cultural background alone would affect performance of students, perceived cognitive load, perceived system usability, perceived ease of use, perceived usefulness, system quality, e-learner's satisfaction, willingness to continue the usage and relative advantage in a summative CBA environment. **H_{5c}**: The combined interacting effect of cultural background and white space or screen density alone would affect performance of students, perceived cognitive load, perceived system usability, perceived ease of use, perceived usefulness, system quality, e-learner's satisfaction, willingness to continue the usage and relative advantage in a summative CBA environment.

3.5.3 Participants and Sample Design

Table 3.3 (next page) gives a snapshot of the participants who took part in the study from the two countries – India and USA. Participants were purposefully classified on the basis of their computer attitude, computer anxiety and previous experiences with CBA systems. Five hundred and sixty two (562) participants volunteered for the study. They were then randomly assigned to one of the two experimental conditions in each of the five experiments. Two hundred and sixty seven (267) students from India and two hundred and ninety five (295) students from USA participated in the experiments. Indian participants were primarily students of the Indian Institute of Technology at Guwahati, Amity University, Delhi and NIT Agartala and the American participants were students from the University of Texas at Austin. One hundred and sixty three (163) male and one hundred

Table 3.3

Participant details

Gender	India	USA
Male	163	121
Female	104	174
n	267	295
Total *n*	562	

and four (104) female students participated from India. One hundred and twenty one (121) male and one hundred and seventy four female students participated from USA. All the participants had experience of taking online computer based tests. Participants were between 18-34 years old (M = 20.76, SD =1.69; India: M=22.71, SD=2.26; USA: M=18.82, SD=1.12). A *between group* design study was formulated for the investigation. While students from India were allowed to choose their location for the experimental study, American students were called in cohorts in a laboratory where each participant was administered a different set of experimental test conditions and no two conditions were similar. This choice for conducting a cohort study in the USA conditions were adopted purely due to limitation of facilities and technical issues. Indian students volunteered for the study while American students were allotted 1 credit hour for their participation.

3.5.4 Instruments Used

In the current research study questionnaires were used as instruments to capture the students' subjective evaluation of the CBA system. Students had to complete the assessment session on computers that comprises of test items and thereafter they were administered the questionnaires. The details of the test items and the questionnaires are listed below.

Test Items

Participants in each condition answered ten (10) questions divided among (i) logical reasoning, (ii) verbal reasoning, and (iii) non-verbal reasoning as delineated in the table 3.4 (next page). The details of the test items are enlisted in **Appendix 1.** Each of the five experiments had different sets of questions. The test conditions within an experiment shared similar questions. Questions were validated with instructional experts regarding the content complexity of the item

Table 3.4

Distribution of test items across the five CBA assessment conditions.

	Exp. No. 1	Exp. No. 2	Exp. No. 3	Exp. No. 4	Exp. No. 5
Logical Reasoning	2	3	1	1	5
Verbal Reasoning	3	2	2	2	4
Non-verbal Reasoning	5	5	7	7	1
Total Questions	10	10	10	10	10

types and after verifying with the experts the test items were prepared. All the five different sets of test items are enlisted in **Appendix-1** shared similar content complexity level among them.

Questionnaires

The study conducted in this investigation employed purposive sampling procedure followed by random selection of participants to the individual test conditions. First, participants were selected on the basis of their computer attitude and computer anxiety. These two factors were captured since attitude toward computer influence students' initial acceptance of IT, and future task behaviour regarding computers (Selwyn, 1997). Similarly computer anxiety is a determinant of perceived ease of use (Gefen, Karahanna, & Straub, 2003a, 2003b; Gefen & Straub, 1997; Pedersen & Nysveen, 2003) and in turn influences the 'likelihood that a student will use' hypermedia based learning system again (Saadé & Kira, 2009). Later, those who were found to have a high computer awareness and low computer anxiety along with previous experiences with CBA systems were randomly allotted the CBA test conditions. The following instruments have been used in the study:

A. Questionnaire to select participants based on their computer awareness and computer anxiety.

i) Learner's attitude towards Computers (Gattiker & Hlavka, 1992).

ii) Learner computer anxiety (Barbeite & Weiss, 2004).

B. After participants were selected with high computer awareness and low computer anxiety, they were randomly directed to the CBA test conditions, which started with a demographic questionnaire before the main assessment session.

Table 3.5

Reliability scores of the scales used in this research study.

Sl. No.	Dependent Measures	*n*	Cronbach's *α*
1.	**Learner's Attitude towards computers**	562	0.77
2.	**Learner Computer Anxiety**	562	0.81
3.	**NASA TLX**	562	0.80
4.	**SUS**	562	0.87
5.	**PEOU**	562	0.82
6.	**PEU**	562	0.90
7.	**SQ**	562	0.82
8.	**RA**	562	0.85
9.	**PES**	562	0.89

C. After the completion of the main assessment session participants were supplied a post-test questionnaire, comprising of the following items:

i. National Aeronautics and Space Administration Task Load Index [NASA TLX (Hart & Staveland, 1988)].
ii. System Usability Scale [SUS (Brooke, 1996)].
iii. Perceived Ease of Use [PEOU (Poelmans, Wessa, Milis, Bloemen & Doom, 2008)].
iv. Perceived Usefulness [PEU (Poelmans, et al., 2008)].
v. System Quality [SQ (Poelmans, et al., 2008)].
vi. Relative advantage [RA (Poelmans, et al., 2008)].
vii. Perceived e-learner satisfaction [PES (Arbaugh, 2000)].
viii. Willingness to continue to use [WCU (Bickmore & Picard, 2004)].

The reliability scores of the scales are presented in table 3.5 (shown above).

3.6 Experimental processes followed

Prior to the main CBA assessment session, participants were administered learner computer awareness questionnaire (Gattiker & Hlavka, 1992), learner computer anxiety questionnaire (Barbeite & Weiss, 2004) and a single item based question to capture their previous experience with CBA systems. Participants thereafter were purposively screened for high learner computer awareness and low computer anxiety and with previous experience with CBA systems. Those, who had scored more than 40 out of 56 in the computer attitude scale and had scored lower than 15 out of 28 in the computer anxiety scale and had previous experiences with CBA systems, were classified as ***expert*** users. After screening the desired participants the CBA system randomly allocated (a random algorithm was used to do the same) the participants a particular CBA test

condition from the repository of the ten CBA test conditions. Screened participants thereafter filled out a demographic questionnaire before entering the main assessment session and questions on their grades (percentage of marks obtained in last qualifying exams). Afterwards participants answered an online computer-based aptitude test that consisted of 10-multiple choice logical, verbal and non-verbal reasoning questions. This test constituted the main part of the experiment (for test items see **Appendix-1**). The graphical representation of the experimental process followed has been depicted in the figure 3.2 (next page).

In the main assessment session ten identical questions consisting of logical reasoning, verbal reasoning, and non-verbal reasoning questions were answered by the participants. They were allotted twenty minutes to answer these questions. In the CBA test condition with customization features two minutes were allotted to the participants for reviewing the questions and carrying out customization accordingly. During the main test session these participants were allotted eighteen minutes to complete the session. The test score and the time taken by each participant to complete the main test were recorded. After completing the online assessment session, participants were administered a post assessment questionnaire to capture their subjective evaluation of the assessment system. Participants were asked to fill out the 7-items NASA Task Load Index [NASA-TLX (Hart & Staveland, 1988) to measure their self-reported cognitive load, the 10-items System Usability Scale (Brooke, 1996), a 4-item scale on perceived ease of use of the CBA (Poelmans, et al., 2008), a 3-item scale on perceived usefulness of the CBA (Poelmans, et al., 2008), a 11-item scale on system quality of the CBA (Poelmans et al., 2008), a 2-item scale assessing the experienced relative advantage of the CBA against traditional PPT (Poelmans et al., 2008), a single item measure of willingness to continue using the CBA (Bickmore & Picard, 2004), as well as a modified version of the 9-item perceived e-learning satisfaction scale (Arbaugh, 2000), in which the word e-learning was changed to e-assessment. All the questionnaires (pre assessment session and post assessment session) were administered to the participants through the online medium.

3.7 Experiment Design

A 2x2 factorial *between group design* was adopted for the study. The study investigates two independent features each with two levels. There factorial design was adopted. No two groups of participants are similar across all the participant groups in the study. Ten dependent variables were captured during the experiments. Therefore, a multivariate analysis technique was adopted for the analysis of the data.

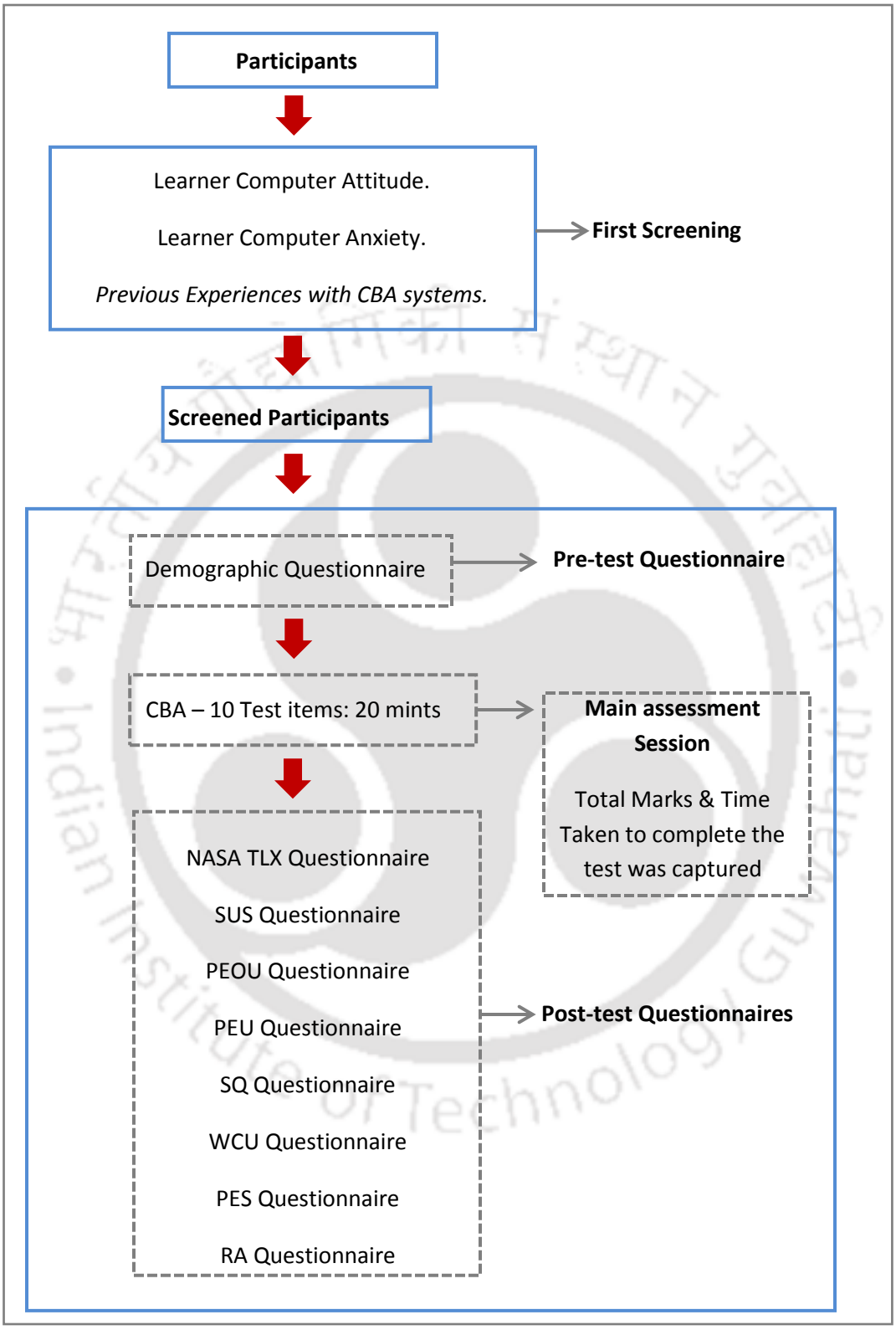


Figure 3.2. Graphical representation of the Experiment Design & Process.

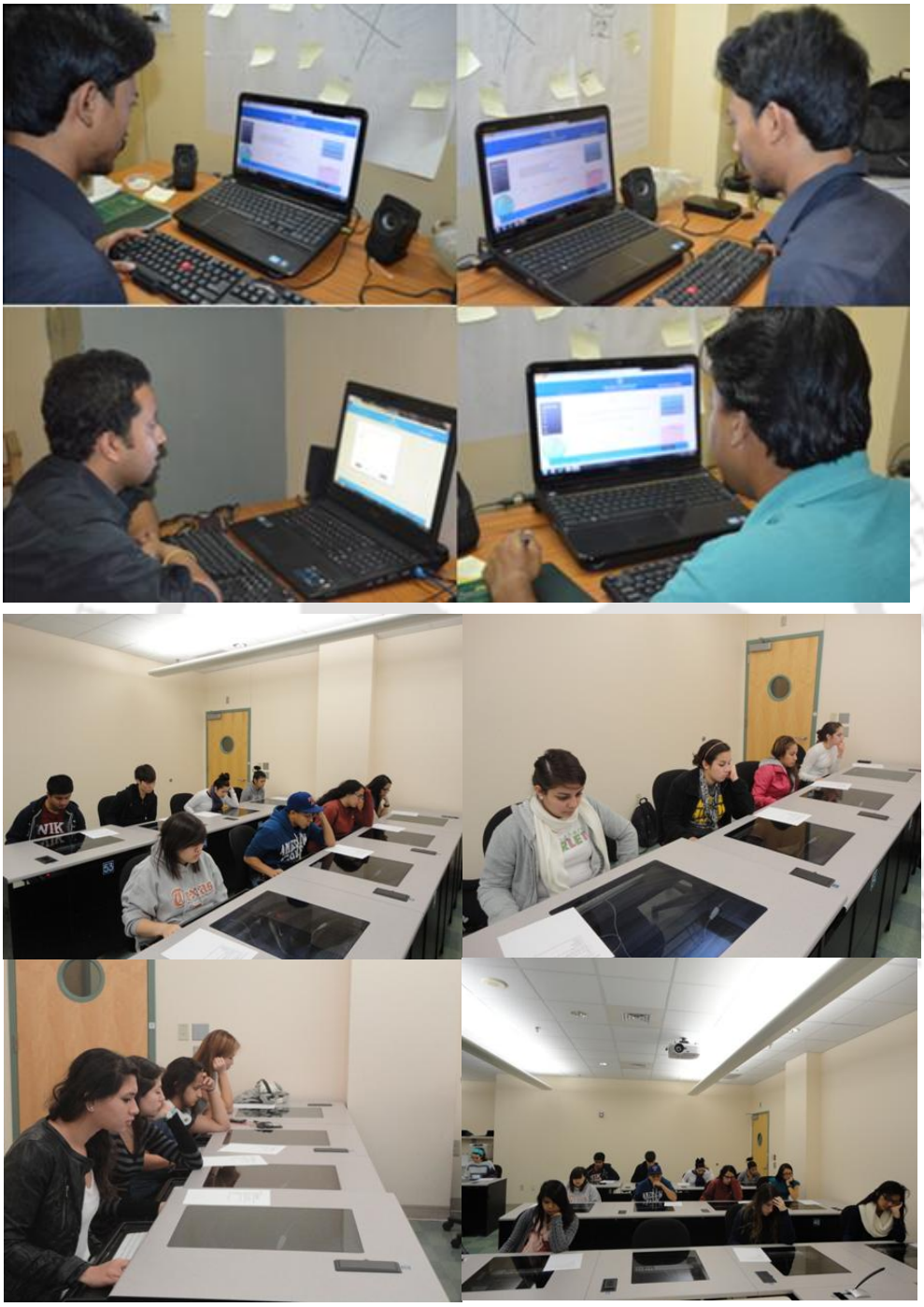


Figure 3.3. Indian & American Participants during the experimental study.

A between groups study has been proposed to reduce learning effects. Dual task experimentation has been used which is ideal to test cognitive load measurements. A graphical illustration of the experimental design is presented in figure 3.4 (next page). Purposive sampling was carried out first to identify students with high computer awareness and low computer anxiety and with previous experiences with CBA systems. After creating a pool of expert students, they were randomly allotted one of the two test conditions out of the five experiments.

3.7.1 Independent Measures

The interface features and cultural backgrounds of the students are the independent factors for the study reported in this thesis. The independent factors for the study with scrolling are - Multiple questions at a time with scrolling and Single Question at a time without scrolling; for navigational mechanisms - Step navigation and vertical menu based navigation; for item review and presentation – a) Single question at a time (question order decided by default by the CBA system) and question customization (Students' can decide the order of the questions for the main assessment session after reviewing the test items) b) question customization and time customization (Students' decide the amount of time to be customized for the main assessment session for each section after reviewing the test items) and for white space (Screen density) – simple single test item with maximum white space and single test item with low white space (white space cluttered with graphical elements). Students from Indian and the United States of America have been identified as the independent factors from the perspective of cultural background for the study.

3.7.2 Dependent Measures

The study reported here captures the performance of the students from the perspective of test scores and time taken to complete the test. But apart from these two variables other sensitive variables such as perceived cognitive load, perceived system usability, perceived ease of use, perceived usefulness, system quality, relative advantage, perceived e-learning satisfaction and willingness to continue to use have also been captured.

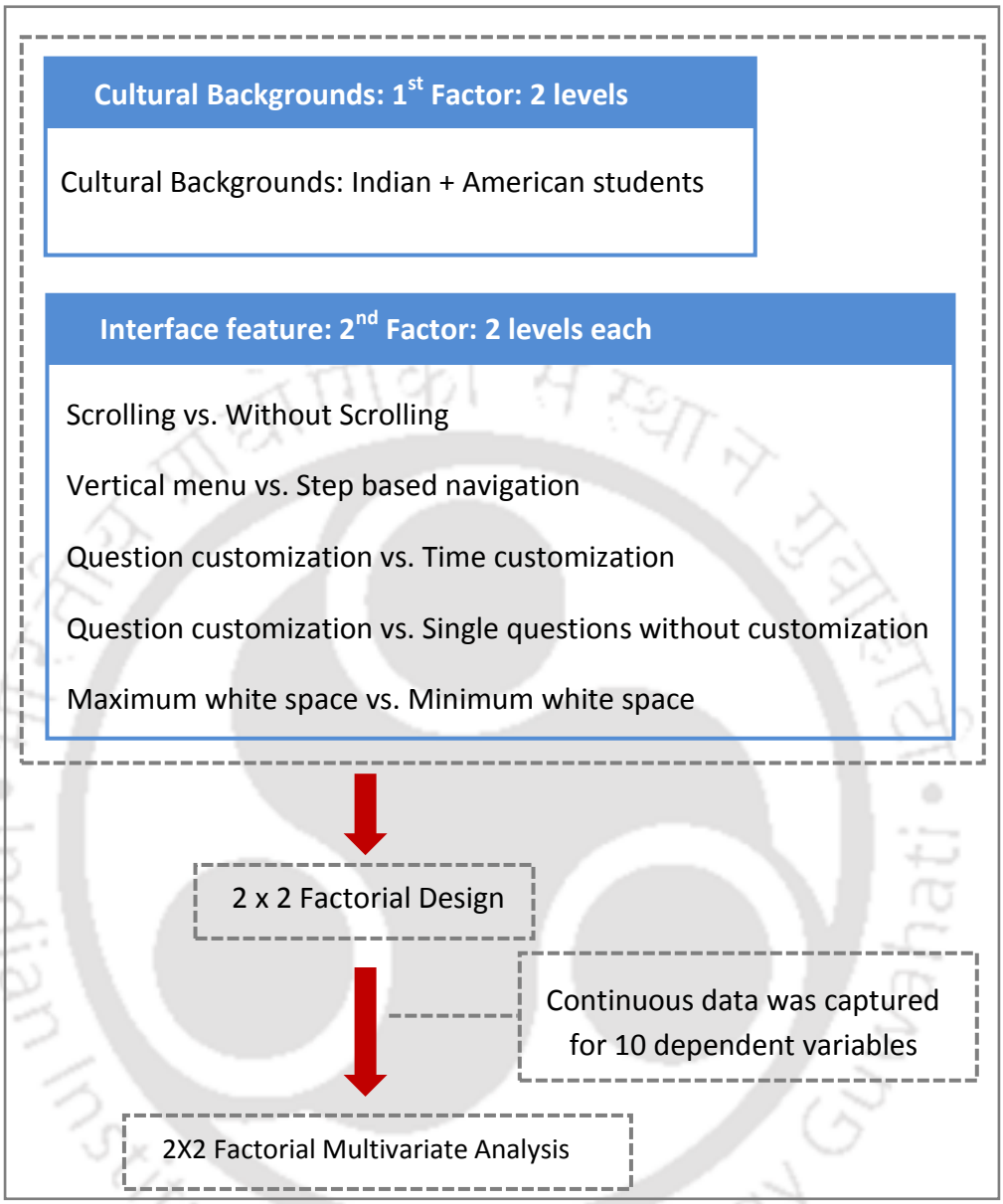


Figure 3.4. Graphical representation of the experimental design approach

Table 3.6

Abbreviations of the dependent measures captured in the studies.

Abbreviations of the dependent variables	
NASA TLX	Perceived cognitive load captured from the NASA Task Load index questionnaire. The dependent measure has been named in this thesis after the questionnaire.
SUS	System Usability Score captured from the System Usability Scale.
PEOU	Perceived Ease of Use.
PEU	Perceived Usefulness.
SQ	System Quality.
RA	Relative Advantage.
PES	Perceived E-learner Satisfaction
WCU	Willingness to Continue to Use.
TM	Total Marks.
TK	Time Taken.

Table 3.6 highlights the abbreviations of the dependent measures captured in the studies. The rationale for choosing these dependent factors is stated below:

i. Perceived Cognitive Load: Cognitive Load affects Performance (Chandler, & Sweller, 1991).
ii. Perceived System Usability & e-learning satisfaction affects success of e-learning system (Sun, Tsai, Finger, Chen & Yeh, 2008)
iii. Perceived Ease of Use and Perceived Usefulness affect Technology Acceptance and thereby affects success of e-learning system (Sun, Tsai, Finger, Chen & Yeh, 2008).
iv. System Quality and E-learner satisfaction affects success of e-learning system (Sun, Tsai, Finger, Chen & Yeh, 2008).
v. Relative advantage captures the preference of students using the current CBA system over other CBA systems.
vi. Willingness to continue to use the system: This factor determines whether the system would be used in the future after initial acceptance.

3.7.3 GUI - Design of Experimental CBA Environment

The screen shots of the CBA system which was specifically designed for the investigations are provided in ten screen shots reproduced in figures 3.5- 3.14. (next page).

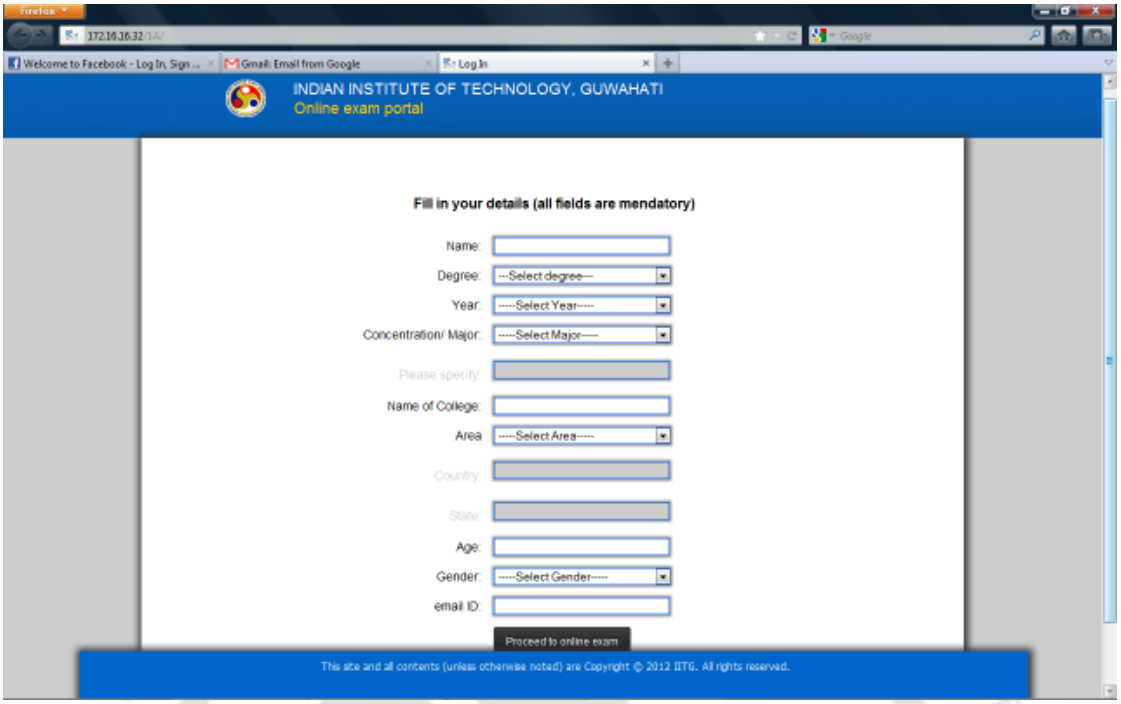


Figure 3.5. Log in screen

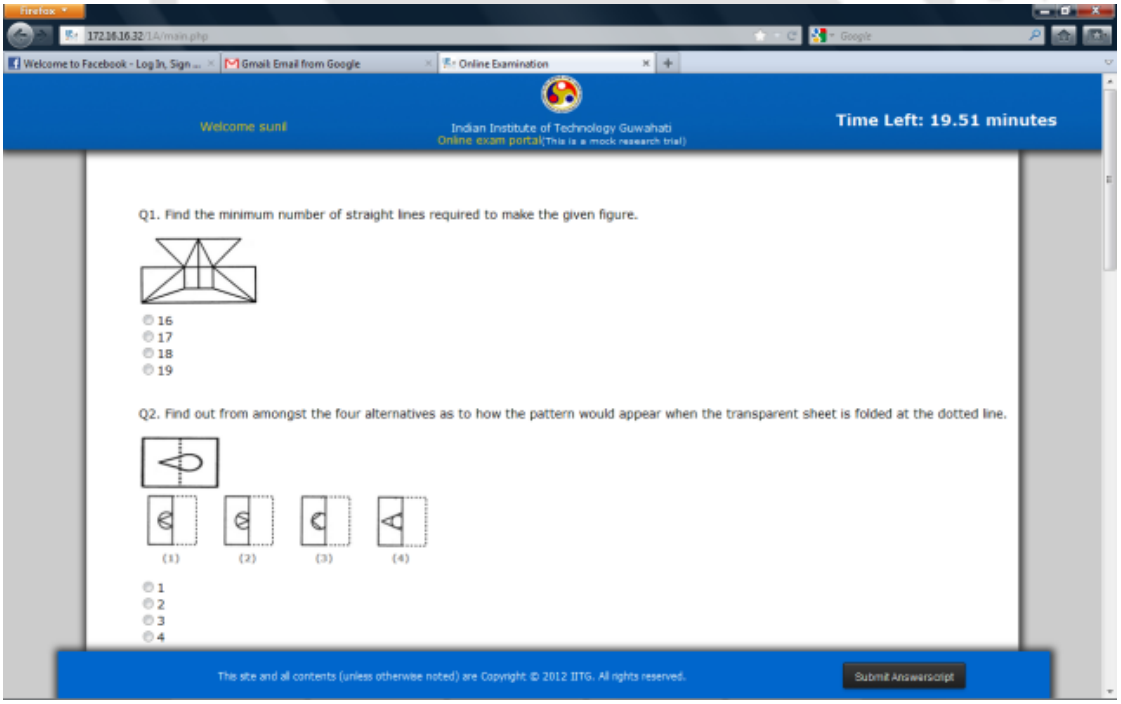


Figure 3.6. Multiple Questions with Scrolling CBA condition.

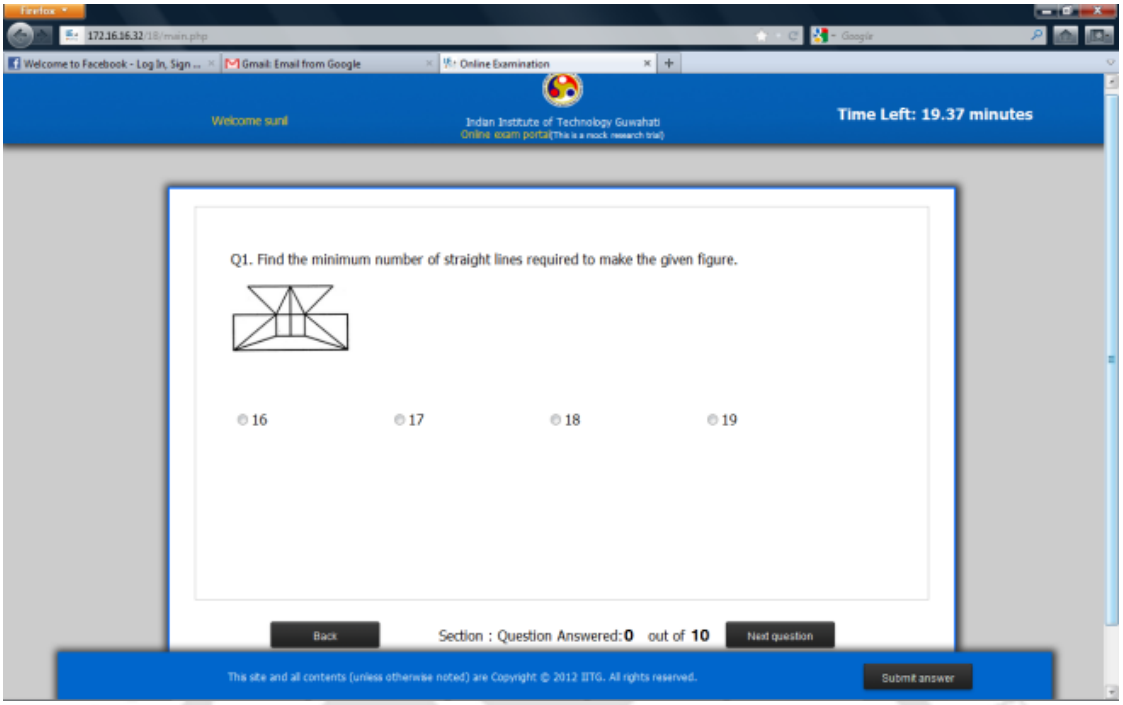


Figure 3.7. Single Question without Scrolling CBA condition.

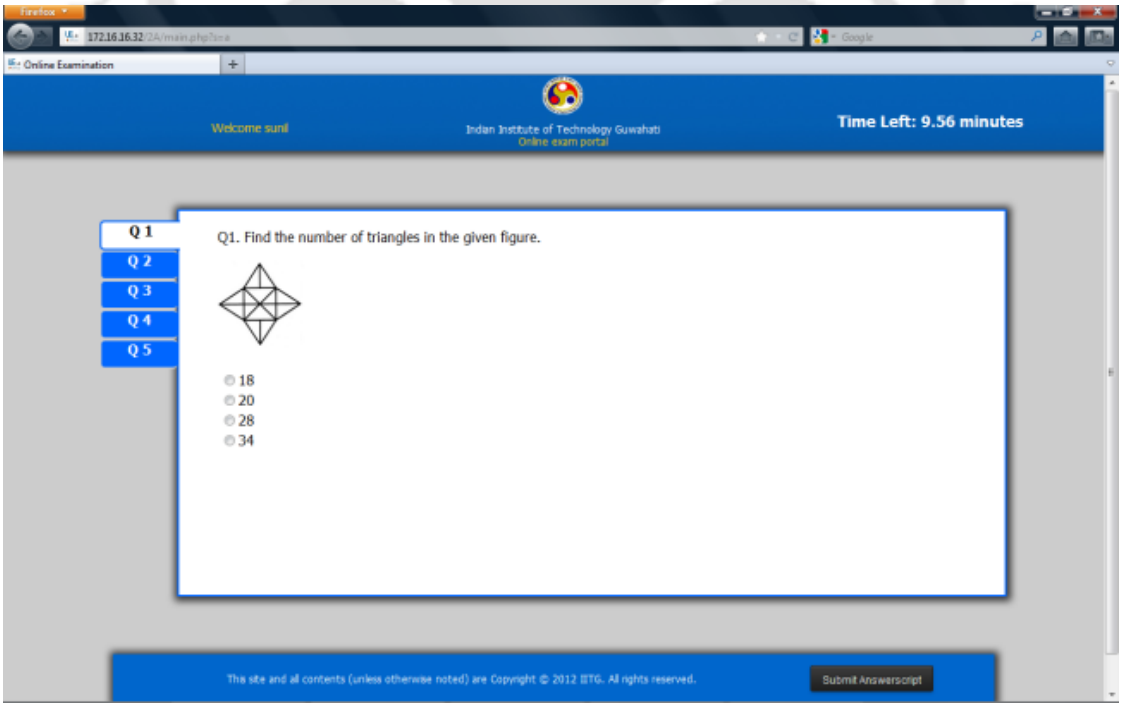


Figure 3.8. Vertical menu based navigation CBA condition.

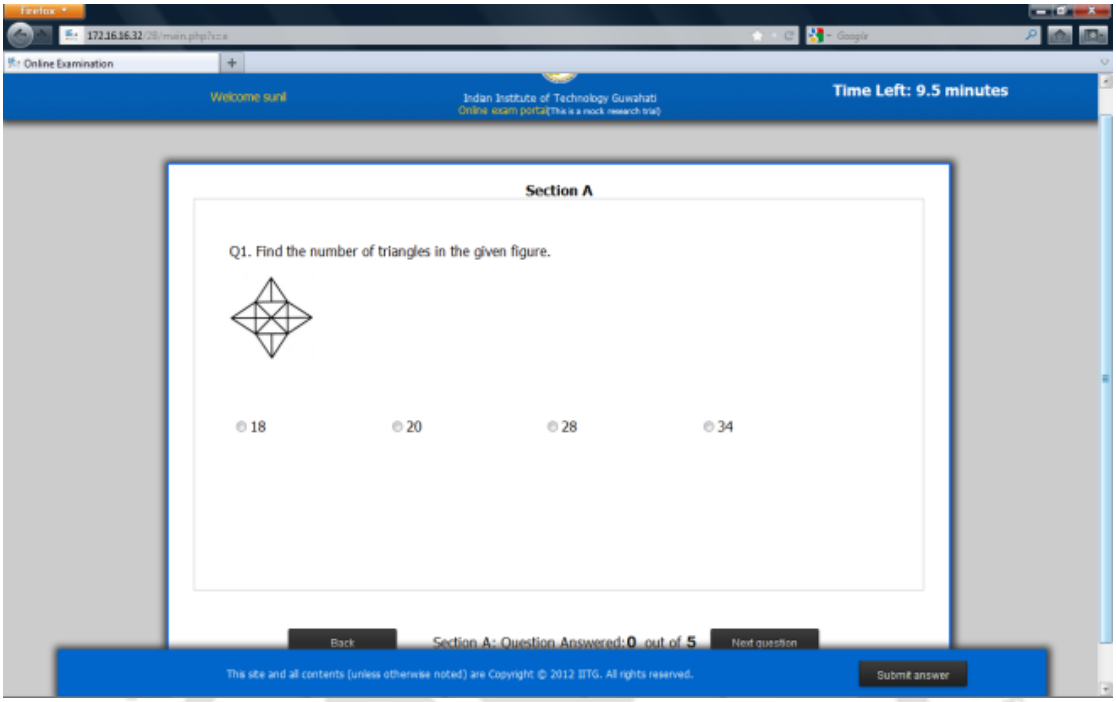


Figure 3.9. Step based Navigation CBA condition.

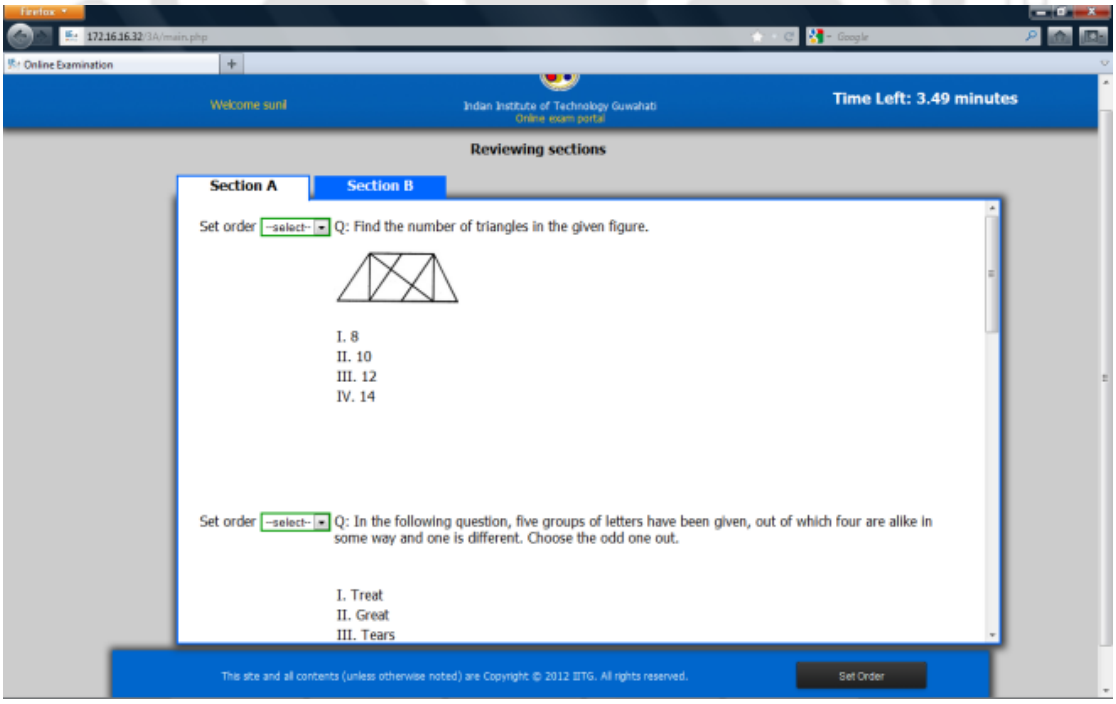


Figure 3.10. Question Customization CBA condition.

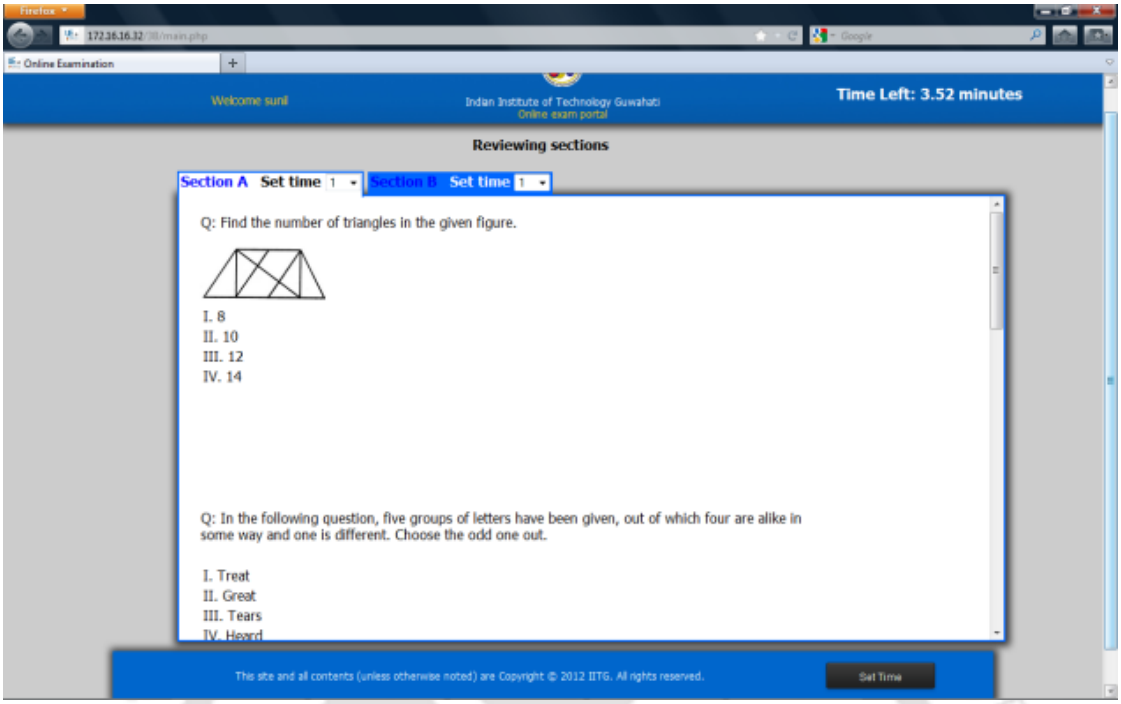


Figure 3.11. Time Customization CBA condition.

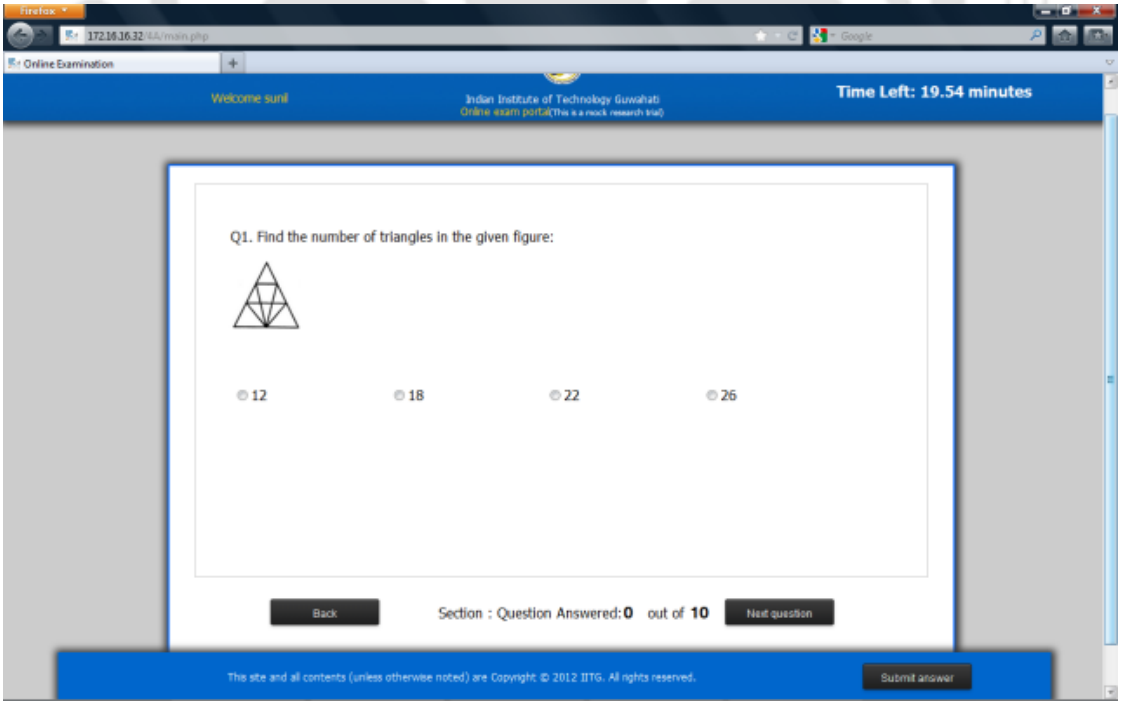


Figure 3.12. Single Question without customization CBA condition

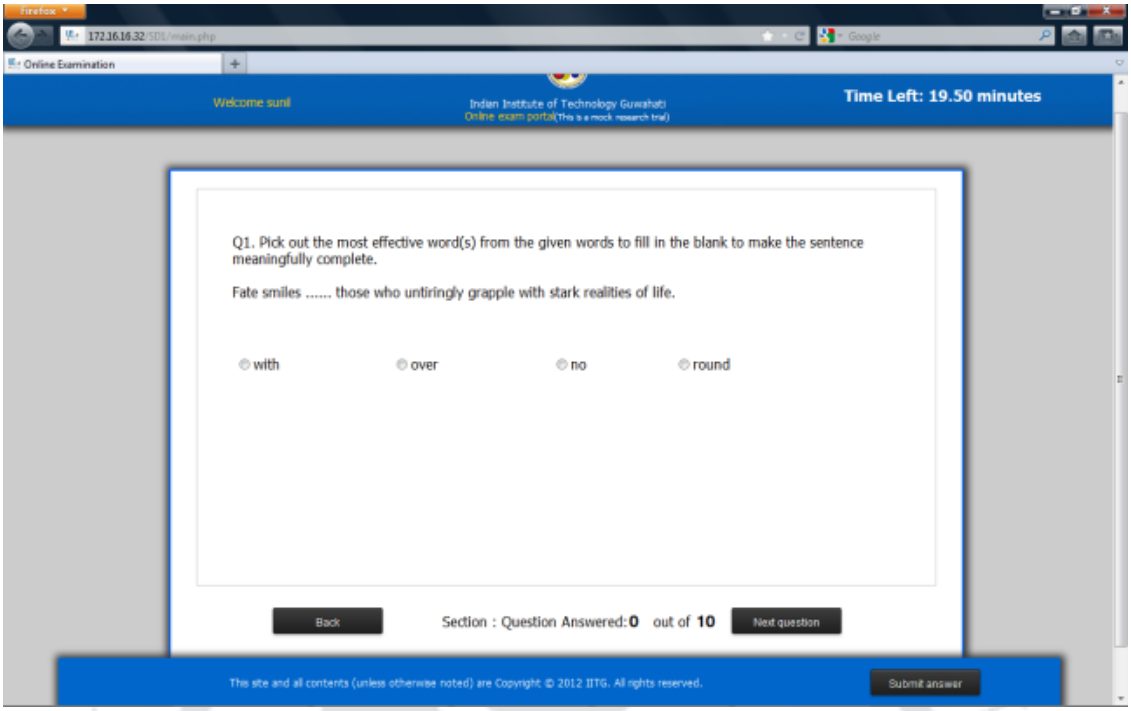


Figure 3.13. Single Question with maximum white space CBA condition

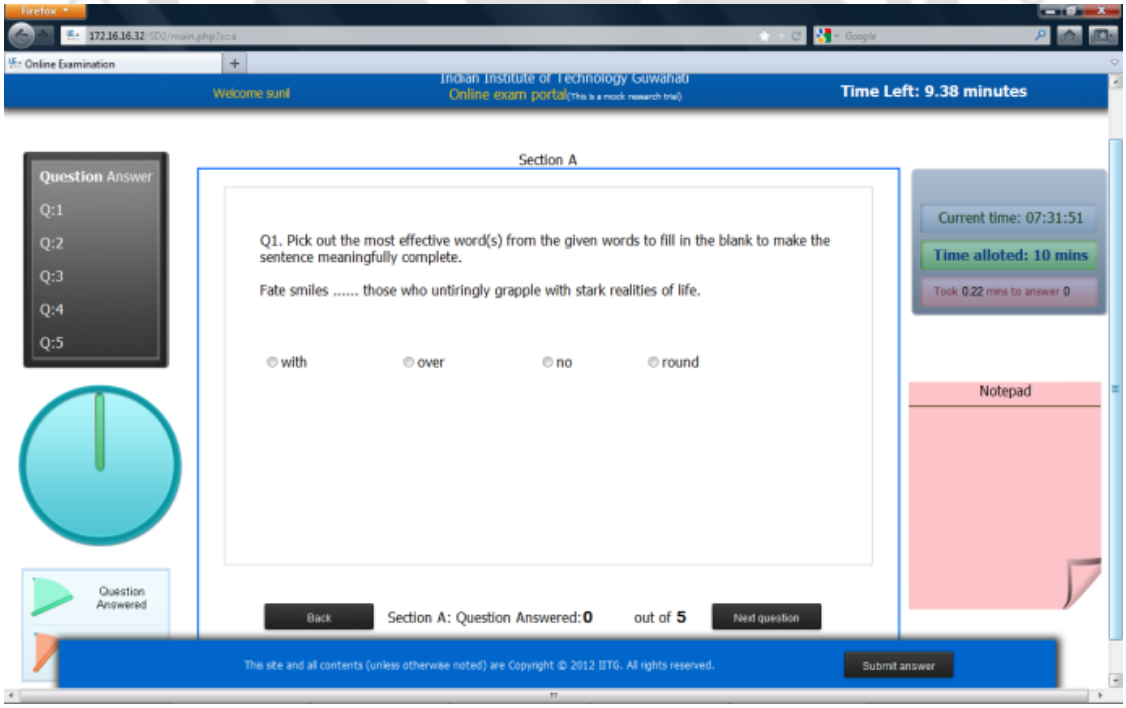


Figure 3.14. Single question with reduced white space CBA condition.

3.8 Experiment Design - Rationale

We adopted a multivariate design for the experiments that were conducted during the research investigation. A multivariate design was primarily adopted because we posited that due to the effect of delivery media performance of students in terms of test scores and their subjective evaluation of the assessment system would get affected. We further posited that even performance of 'Experts' will be affected because of the effect of delivery media. Here, expert users and novice users were as defined by Prümper, Frese, Zapf and Brodbeck, "All in all, there are usually two (strongly overlapping) criteria used for the differentiation between novices and experts: knowledge (e .g. comparing students and teachers) and the time spent working with a particular system (e .g. students with a few vs. those with many courses)" (Prümper, Frese, Zapf & Brodbeck, 1991). During the experiments, we classified the students on the basis of their computer attitude and their computer anxiety. Those students' who reported higher computer attitude, lower computer anxiety and had previous experiences with CBA systems (more than once) and had undertaken courses or exams previously as a part of their curriculum program have been classified as expert users.

Performance in test mode effect studies have been captured primarily from the students test scores. But, literature has also suggested that cognitive load might also act as a predictor of students' performance in computer based assessment environment. Sharon Oviatt (2006) states that, "Cognitive Load Theory (CLT) and its cognitive science underpinnings provide a coherent and powerful basis for predicting performance when using alternative interfaces, and for designing interfaces that effectively minimize cognitive load (Baddeley, 1992; Baddeley, 2003; Mousavi, Low & Sweller, 1995; Oviatt, Arthur & Cohen 2006; Paas, Tuovinen, Tabbers & Van Gerven, 2003; Sweller, 1988; Tindall-Ford, Chandler & Sweller, 1997; Van Merrienboer & Sweller, 2005; Wickens, 2002; Wickens, Sandry & Vidulich, 1983)" (Oviatt, 2006). This argument is also coherent from the perspective of learner centered approach. While, users in a User centered design approach (UCD) are considered to be experts who are proficient in their work activities; in learner centered approach (LCD) the users are trying to learn their new work activities through the system, gradually moving from novice users to experts.

In order, to achieve an efficient hypermedia based assessment which is an intrinsic component of the hypermedia based learning system it is important that the system should help the students to spend more time on completing the task (task refers to the assessment module) to achieve their learning objectives (in this context the learning objective refers to the successful completion of answering the test items) rather than forcing them to spend more time in learning the tool itself (the CBA system). Cognitive load which is powerful basis for predicting performance

was therefore chosen to be investigated in the context of the current CBA system. In a CBA environment where mastering difficult interface is a form of problem solving, some individuals adapt and learn quickly while others take more time to build their mental model of the interface (Drommi, Ulferts & Shoemaker, 2001). Therefore, satisfaction of students using the CBA system is an important indicator that determines the success of information systems (Delon & Mclean, 1992) in efficiently assisting the students to spend time and cognitive resources to complete the task rather than forcing them to focus on learning the tool (in this case the interface features) to complete the task. Usability tests are generally used to measure the effectiveness of interfaces (Cheon & Grant, 2009). So, usability, perceived quality and satisfaction of the presentation formats were captured from the participants by using questionnaires.

In each of the five experiments ten dependent variables were captured. Interface features and cultural backgrounds of the participants were the independent measures. Each of the two independent measures had two levels. Therefore, in this research investigation a *multivariate design* has been adopted.

3.9 Preparing for Data Analysis

Data from the post-test questionnaire were continuous in nature. The experiment was conducted to capture the effect of the interface features and cultural backgrounds of the participants on the ten dependent variables namely, test score, time taken to complete the test, system usability score, perceived cognitive load, perceived ease of use, perceived usefulness, perceived e-learner satisfaction, willingness to continue to use, relative advantage, system quality. Since both the independent variables namely the interface features and cultural background has two levels, therefore a 2x2 factorial multivariate analysis were adopted to analyse the data. First, the entire data set was tested for normality, thereafter it was tested for the assumption of within group homogeneity of variance- covariance matrices. There are the basic assumptions for carrying out the multivariate analysis MANOVA. Pillai's trace statistic was reported in the multivariate analysis as it is the most serviceable of the four test statistic namely, Wilk's, Hotelling's and Roy's criterion because of its robustness under violations of the assumptions (Haase & Ellis, 1987).

The hypotheses raised in this research study were exploratory in nature. This investigation intends to study the effects of the treatments on several criterion variables individually. Therefore to control for Type 1 error first a MANOVA was employed and thereafter separate univariate analysis are performed. This approach of analysis of separate follow-up univariate tests after a significant

MANOVA has been highlighted by Bray and Maxwell in 1982. The study restricts its scope to the identification of effects of the treatment groups on the criterion variables. Therefore, the intercorrelations among the criterion variables were not taken into account during the analysis though they have been reported. Generally, different follow-up tests are performed that takes into consideration the intercorrelations of the criterion variables, but such analysis are followed only when the investigator is interested in the relationships among the criterion variates or the investigator wishes to reduce the criterion variates to some smaller set of psychological or theoretical dimensions or the investigator is interested in the set of measures as they represent some underlying construct or dimension (Bray & Maxwell, 1982). This investigation addresses issues from the perspective of user interface design approach for CBA systems. As such the study is more focused on investigating the design issues (here the effect of the interface features) rather studying the combination or the contribution of the criterion variates.

The data collected during the Experiments have been analysed statistically and presented in the next Chapter.

3. 10 Chapter 3: Summary

A detailed description of the research methodology of the study under investigation has been reported in this chapter. Working hypotheses of the five experiments conducted were presented. Instruments used for data collection, nature of the participants recruited during the study and experiment design has been discussed in detail. Justification of the multivariate analysis procedure adopted had been stated. The statistical analysis techniques employed for data analysis presented in chapter 4 has been thoroughly discussed with rationales in this chapter. Data analysis and discussion have been presented in the next chapter.

Chapter 4

Results, Analysis and Discussions

Chapter Abstract: *This chapter reports the statistical analysis of the data collected during the five experimental studies. A 2x2 factorial multivariate analysis approach has been adopted to analyse the data sets. Each experimental result highlights the effect of the corresponding interface feature and cultural backgrounds of the students on the performance and their subjective evaluation of the CBA system. Detailed discussion of the results from the perspective of the working hypotheses of the experiments has been done.*

4.1. Introduction

The research study reported in this thesis investigates the effect of interface features and cultural backgrounds of the students on their performance and their subjective evaluation of the assessment mode in a CBA system. It is important to carry out such an investigation because literature fails to address the issue of the impact of interface features and cultural background on "test mode effect". Though the current study does not investigate "test mode effect" per se but it investigates the effect of various presentation formats within the same mode, here the computer based mode of assessment delivery. Five experiments were conducted with students from India and the United States of America. Participants were purposefully screened for higher learner computer attitude, low computer anxiety and with previous experience with CBA systems. Each of the five experiments had five different interface features as treatments. Treatments like scrolling, navigation, customization and white space were manipulated in each experiment. The effect of these interface features and cultural backgrounds were observed on the dependent measures namely, subjective cognitive load, perceived system usability, perceived ease of use, perceived usefulness, perceived system quality, relative advantage, willingness to continue to use, perceived e-learner's satisfaction, total marks obtained in the test and time taken to complete the test.

The current chapter presents the statistical analysis of the data collected during the five experimental studies. Working hypotheses of the five experiments have been stated again and the results were disused in the light of these hypotheses.

4.2. Results and Analysis

A 2x2 factorial *multivariate analysis* technique was adopted for analysing the data. It was followed by *multiple univariate tests* (ANOVA) if the results of MANOVA were found significant. First, tests of normality was carried out, thereafter the assumption of homogeneity of within group variance- covariance matrices were investigated through the Box's M test. After the tests of the assumptions, MANOVA was employed to analyse the multivariate data sets.

Once a significant MANOVA is observed it was followed again by Levene's test for homogeneity of variances at each individual dependent measure. Thereafter, Univariate Tests were employed to identify the constructs that were responsible for significant MANOVA effect.

The subsequent sections in this chapter enlist the detailed results of each of the five experiments. Working hypotheses for each of the five experiments are presented and subsequent analysis is reported.

4.2.1 Experiment 1: Scrolling vs. Without Scrolling

The working hypotheses which were tested in the experiment scrolling vs. without scrolling are listed below:

H_{1a}: Scrolling alone would affect performance of students, perceived cognitive load, perceived system usability, perceived ease of use, perceived usefulness, system quality, e-learner's satisfaction, willingness to continue the usage and relative advantage in a summative CBA environment.

H_{1b}: Cultural background alone would affect performance of students, perceived cognitive load, perceived system usability, perceived ease of use, perceived usefulness, system quality, e-learner's satisfaction, willingness to continue the usage and relative advantage in a summative CBA environment.

H_{1c}: The interacting effect of cultural background and scrolling alone would affect performance of students, perceived cognitive load, perceived system usability, perceived ease of use, perceived usefulness, system quality, e-learner's satisfaction, willingness to continue the usage and relative advantage in a summative CBA environment.

Table 4.1 and table 4.2 highlight the descriptive statistics of the participants and the dependent measures that were captured during the experiment. In the *with scroll* test condition 27 students (16 male and 11 female; M_{Age} = 23.7, SD_{Age} = 2.13; M_{Marks} = 74.62, SD_{Marks} = 12.38) from India participated, while 26 students (11 male and 15 female; M_{Age} = 19.03, SD_{Age} = 1.53; M_{Marks} = 72.92, SD_{Marks} = 10.45) from USA participated. For the *without scroll* test condition again 27 students (16 male and 11 female; M_{Age} = 21.92, SD_{Age} = 1.54; M_{Marks} = 73.92, SD_{Marks} = 12.27) from India participated, while 29 students (14 male and 15 female; M_{Age} = 18.72, SD_{Age} = 0.92; M_{Marks} = 73.96, SD_{Marks} = 11.20) from USA participated. Groups were homogeneous in terms of their marks obtained in the last qualifying exam but not in terms of their age (See Appendix 4, page 240).

Table 4.1

Descriptive statistics of participant characteristics from both the countries.

	With Scroll Condition		**Without Scroll Condition**	
	India	USA	India	USA
Gender	Male: 16 Female: 11 Total: 27	Male: 11 Female: 15 Total: 26	Male: 16 Female: 11 Total: 27	Male: 14 Female: 15 Total: 29
Age	23.37 (*M*) 2.13 (*SD*)	19.03 (*M*) 1.53 (*SD*)	21.92 (*M*) 1.54 (*SD*)	18.72 (*M*) 0.92 (*SD*)
Marks obtained in last qualifying exam	74.62 (*M*) 12.38 (*SD*)	72.92 (*M*) 10.45 (*SD*)	73.92 (*M*) 12.27 (*SD*)	73.96 (*M*) 11.20 (*SD*)

Table 4.2

Descriptive statistics of the dependent measures.

	With Scroll				**Without Scroll**			
	India		**USA**		**India**		**USA**	
	M	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***
NASATLX	41.24	13.77	35.83	12.54	26.65	14.88	39.22	16.47
SUS	75.41	12.09	85.00	12.33	82.41	14.13	77.84	15.36
PEOU	14.67	2.76	17.15	3.52	17.48	2.85	16.62	4.50
PEU	10.33	2.51	12.88	2.90	11.81	2.77	12.24	3.49
SQ	39.48	5.11	45.73	7.31	44.63	6.87	44.79	6.13
RA	6.22	1.63	6.00	2.08	7.15	1.78	6.45	1.94
PES	43.67	8.56	47.54	13.86	50.44	9.62	45.07	12.19
WCU	6.63	1.36	7.04	2.51	8.00	1.80	7.86	1.79
TM	6.41	1.58	5.92	1.38	6.00	1.66	6.21	1.54
TK	14.38	3.54	8.20	2.55	12.86	4.12	9.64	3.05

Table 4.3

Summarized results of the tests of Normality (Q-Q plots).

Dependent Measures	With Scrolling		Without Scrolling	
	India	**USA**	**India**	**USA**
NASA TLX	Almost Normal	Almost Normal	Almost Normal	Almost Normal
SUS	Almost Normal	Slight deviations	Slight deviations	Slight deviations
PEOU	Perfectly Normal	Almost Normal	Slight deviations	Slight deviations
PEU	Perfectly Normal	Almost Normal	Slight deviations	Slight deviations
SQ	Slight deviations	Almost Normal	Slight deviations	Almost Normal
RA	Almost Normal	Almost Normal	Almost Normal	Almost Normal
WCU	Almost Normal	Almost Normal	Slight deviations	Slight deviations
PES	Almost Normal	Almost Normal	Almost Normal	Slight deviations
TM	Perfectly Normal	Perfectly Normal	Perfectly Normal	Perfectly Normal
TK	Almost Normal	Almost Normal	Almost Normal	Slight deviations

The inferential statistical analysis was carried out first starting with the tests of Normality. Q-Q plots were generated to test the normality of the data distribution across the groups for all the dependent measures. Appendix-3A enlists the Q-Q plots of the data distributions. The observations are summarized in table 4.3. Though the Q-Q plots revealed that there are slight deviations from Normality for some of the dependent measures but the Pillai's statistic is robust to the violations of these assumptions (Haase & Ellis, 1987). Thereafter, the assumption of homogeneity of variance-covariances were tested using Box's Test. Box's test of equality of covariance matrices highlights that it is not significant at $p>0.005$. Therefore, the assumption of homogeneity of variance-covariance matrices was not violated.

The 2x2 factorial multivariate tests (MANOVA) conducted on the data sets revealed (Table 4.4) that there was a statistically significant main effect for interface feature, F (10, 95) = 1.92, $p < 0.05$; Pillai's Trace = 0.16; partial $\eta^2 = 0.16$. There was also a statistically significant main effect for Country, F (10, 95) = 9.07, $p < 0.001$; Pillai's Trace = 0.48; partial $\eta^2 = 0.48$. There was also a statistically significant interaction effect for Country and interface feature when combined together, F (10, 95) = 2.45, $p < 0.05$; Pillai's Trace = 0.20; partial $\eta^2 = 0.20$. It is important to state here that because of the observed significant interaction effect between country and interface, the main effects of these criterion variables which have been found significant no longer demands importance. It is evident from the significant interaction effect that one level of the criterion variable depends on the level of the other criterion variable to affect the group measures. Therefore, after the significant interaction MANOVA effect, separate univariate tests (Table 4.5) was carried out to figure out the dependent measures which account for significant differences across the groups.

Table 4.4

Multivariate Tests

Effect		Value	*F*	Hypothesis df	Error df	Sig.	Partial Eta Squared	Observed Power[c]
Exam	**Pillai's Trace**	0.16	1.92[b]	10.00	95.00	.048	0.16	0.83
Country	**Pillai's Trace**	0.48	9.07[b]	10.00	95.00	.00	0.48	1.00
Exam * Country	**Pillai's Trace**	0.20	2.45[b]	10.00	95.00	.01	0.20	0.92

Table 4.5

Univariate tests for the interaction effects between Exam and country.

Source	Dependent Variable	Type III Sum of Squares	df	Mean Sq.	*F*	Sig.	Partial Eta Squared	Observed Power[k]
Exam * Country	**NASATLX**	2255.14	1	2255.14	10.58	0.00	.09	0.89
	SUS	1312.70	1	1312.70	7.04	0.00	.06	0.74
	PEOU	71.20[a]	1	71.20	5.79	0.01	.05	0.66
	PEU	28.05[b]	1	28.05	3.20	0.07	.03	0.42
	SQ	226.80[c]	1	226.80	5.56	0.02	.05	0.64
	RA	1.57[d]	1	1.57	.45	0.50	.00	0.10
	PES	578.19[e]	1	578.19	4.52	0.03	.04	0.55
	WCU	1.48[f]	1	1.48	0.41	0.52	.00	0.09
	TM	3.97[g]	1	3.97	1.67	0.19	.01	0.24
	TK	65.56[h]	1	65.56	5.82	0.01	.05	0.66

Univariate analysis conducted reveals that the group membership, the interacting criterion variables interface feature and cultural background had a significant effect on subjective cognitive load, $F(1, 104) = 10.58$, $p < 0.05$; partial $\eta^2 = 0.09$; perceived system usability, $F(1, 104) = 7.04$, $p < 0.05$; partial $\eta^2 = 0.06$; perceived ease of use, $F(1, 104) = 5.79$, $p < 0.05$; partial $\eta^2 = 0.05$; perceived system quality $F(1, 104) = 5.56$, $p < 0.05$; partial $\eta^2 = 0.05$; perceived e-learner satisfaction $F(1, 104) = 4.52$, $p < 0.05$; partial $\eta^2 = 0.04$; time taken, $F(1, 104) = 5.82$, $p < 0.05$; partial $\eta^2 = 0.05$. Though, it was clear from the separate univariate tests that group differences among subjective cognitive load, perceived system usability, perceived

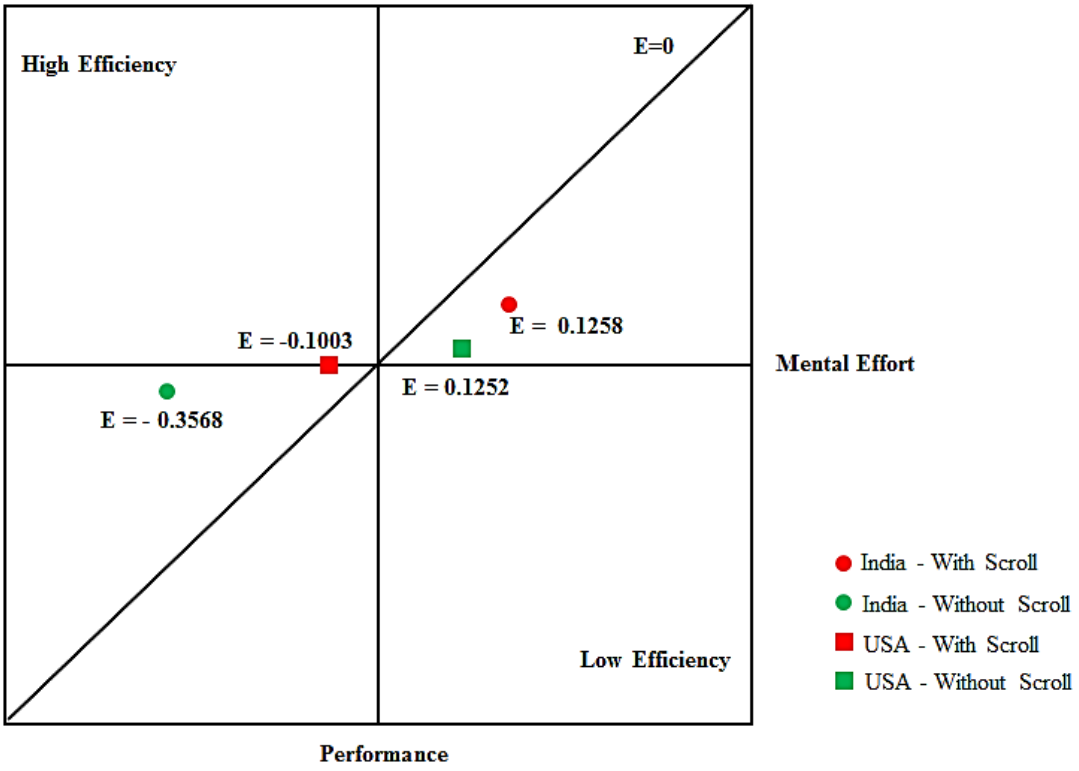


Figure 4.1. Instructional efficiency of 'with scroll' and 'without scroll' CBA conditions between Indian and American students.

ease of use, perceived system quality, perceived e-learner's satisfaction and time taken can be explained by the interacting criterion variables but further analysis is required to identify the particular groups that differ significantly.

To identify the significant differences among groups separate independent sample *t*-tests were employed. **Appendix- 3A** enlists the independent sample *t*-test tables. It is important to state here that *t*-tests are generally robust against violations of normality (Edgell & Noon, 1984). After *t*-tests were conducted on *with scroll* condition among the Indian and the American students it revealed that there was a significant difference in the perceived system usability scores in the *with scroll* condition for Indian students (M = 75.40, SD = 12.08) and American students (M = 85.00, SD = 12.32); t (51) = -2.86, p = 0.006, d= -0.8. A significant difference in the perceived ease of use scores in the with scroll condition for Indian students (M = 14.66, SD = 2.75) and American students (M = 17.15, SD = 3.51); t (51) = -2.87, p = 0.006, d= -0.8. A significant difference in the perceived system quality scores in the *with scroll* condition for

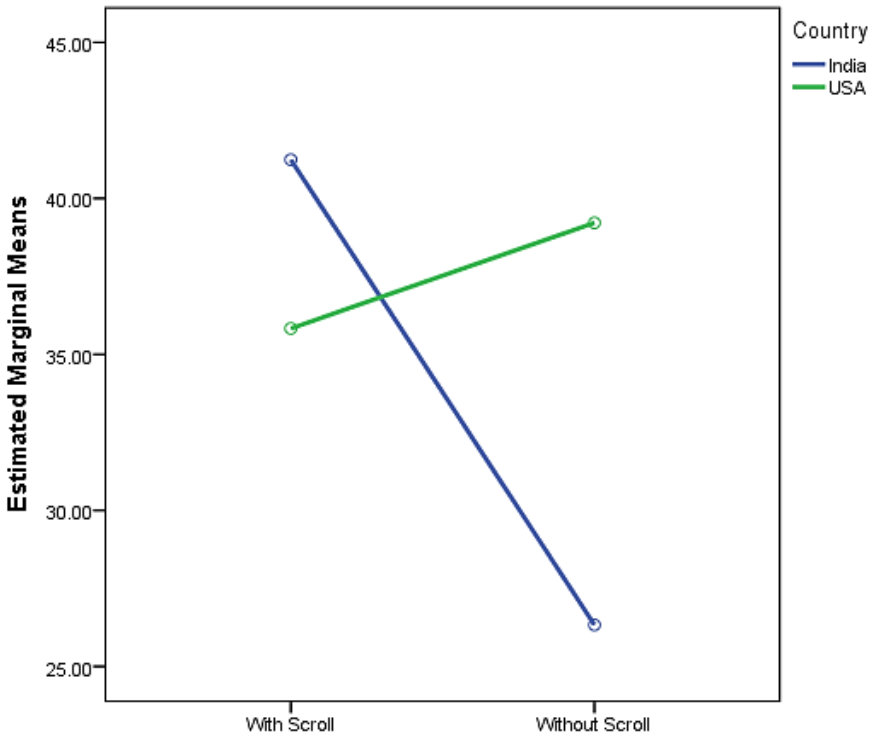


Figure 4.2. Estimated Marginal means of NASA TLX across the test conditions and countries.

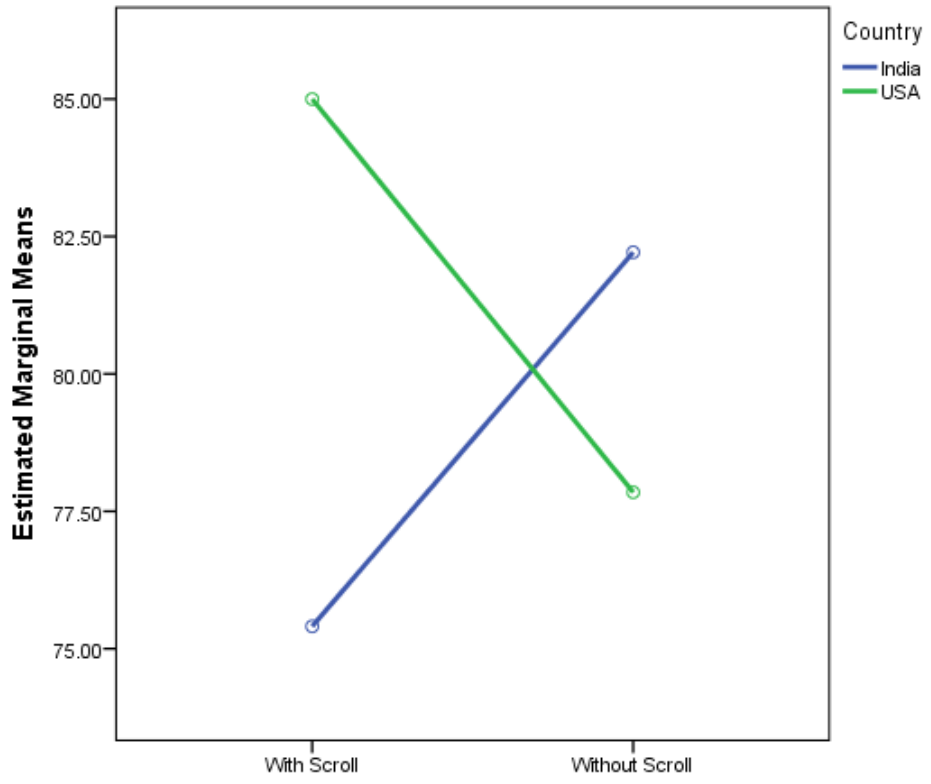


Figure 4.3. Estimated Marginal means of Perceived system usability across the test conditions and countries.

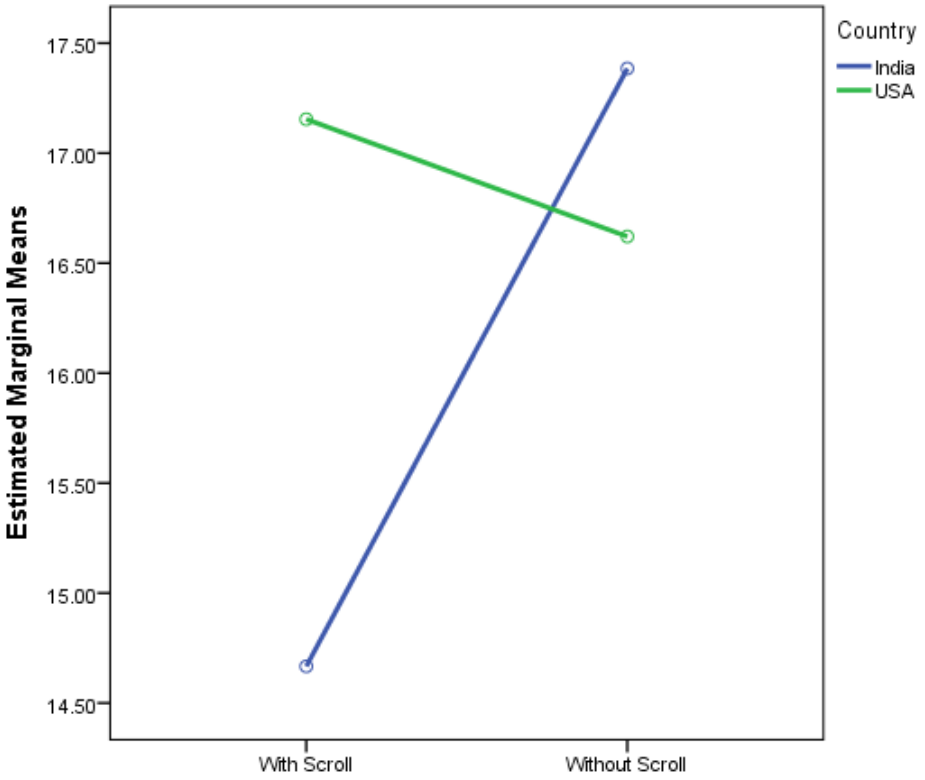


Figure 4.4. Estimated Marginal means of Perceived ease of use across the test conditions and countries.

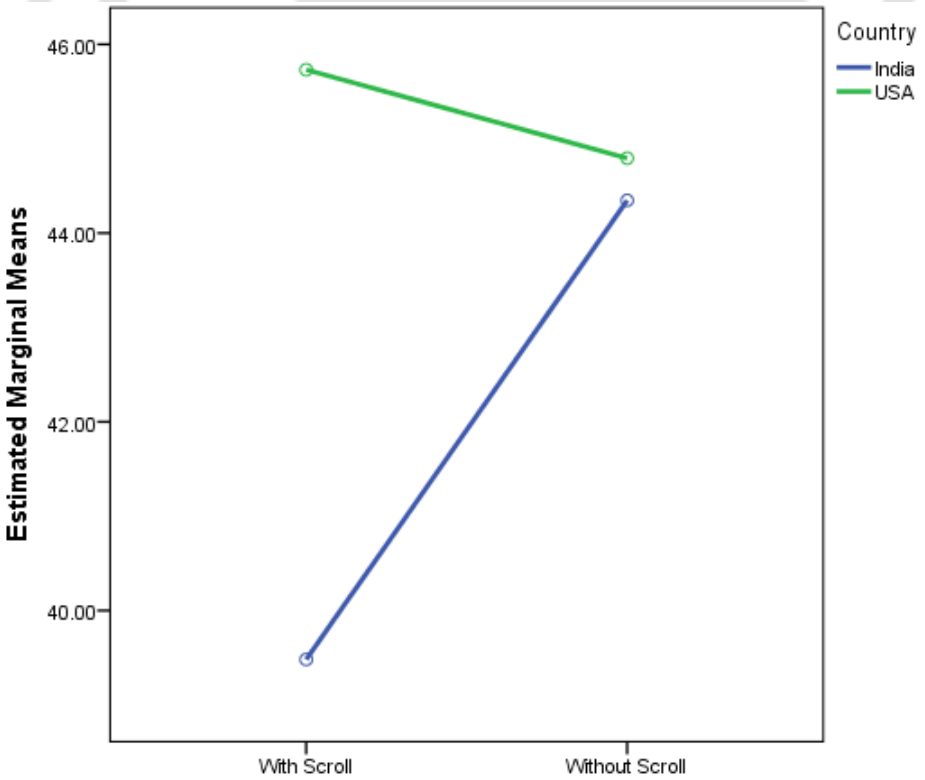


Figure 4.5. Estimated Marginal means of Perceived system quality across the test conditions and countries.

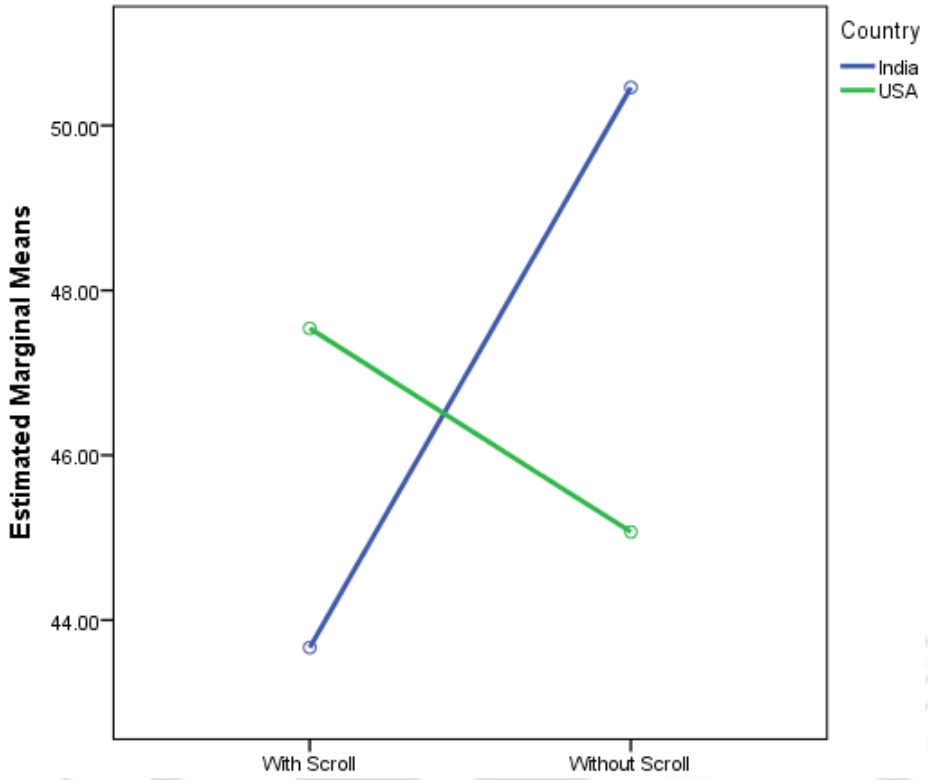


Figure 4.6. Estimated Marginal means of Perceived e-learner's satisfaction across the test conditions and countries.

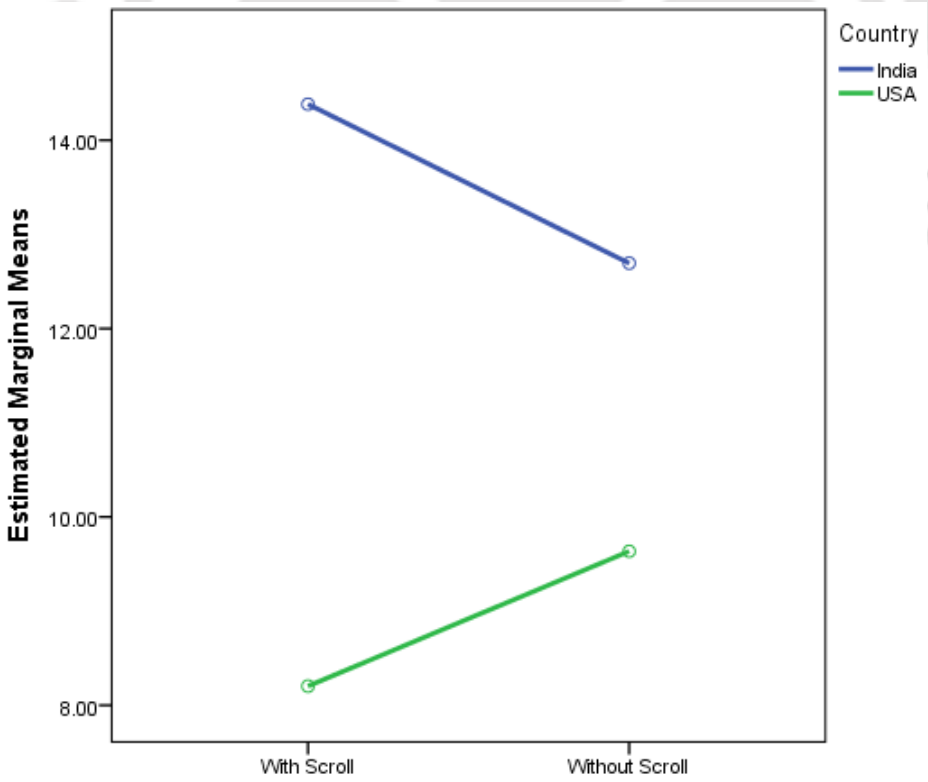


Figure 4.7. Estimated Marginal means of time taken across the test conditions and countries.

Table 4.6

Hypotheses test results

Hypothesis $\mathbf{H_{1c}}$ = Cultural background X Interface feature affects	Supported
Subjective Cognitive Load	Supported
Perceived System Usability	Supported
Perceived Ease of Use	Supported
Perceived Usefulness	Not supported
Perceived System Quality	Supported
Relative Advantage	Not supported
Perceived E-learner's Satisfaction	Supported
Willingness to Continue to Use	Not supported
Total Marks	Not supported
Time Taken	Supported

Indian students (*M* = 39.48, *SD* = 5.10) and American students (*M* = 45.73, *SD* = 7.31); *t* (51) = -3.618, *p* = 0.001, *d*= -1.01. A significant difference in the time taken scores in the *with scroll* condition for Indian students (*M* =14.38, *SD* =3.54) and American students (*M* =8.20, *SD* =2.54); *t* (51) =7.26, *p* =0.000, *d*= 2.03.

The *t*-tests when employed for the *without scroll* condition revealed a significant difference in the subjective cognitive load scores for Indian students (*M* =26.64, *SD* =14.88) and American students (*M* =39.21, *SD* =16.47); *t* (54) = - 2.98, *p* =0.004, *d*= -0.81. A significant difference in the time taken scores for Indian students (*M* = 12.85, *SD* =4.11) and American students (*M* =9.63, *SD* =3.04); *t* (54) =3.34, *p* = 0.002, *d*= 0.9.

After, conducting individual sample *t*-tests for the interface conditions, it was again employed for the cultural background conditions. The *t*-tests when employed for the Indian students revealed a significant difference in the subjective cognitive load scores for *with scroll* condition (*M* = 41.24, *SD* = 13.77) and *without scroll* condition (*M* =26.64, *SD* =14.88); *t* (52) = 3.74, *p* =0.000, *d*= 1.03. A significant difference in the perceived ease of use scores for *with scroll* condition (*M* =14.66, *SD* =2.75) and *without scroll* condition (*M* =17.48, *SD* =2.84); *t* (52) = -3.68, *p* =0.001, *d*= -1.02. A significant difference in the perceived system quality scores for *with scroll* condition (*M* = 39.48, *SD* = 5.10) and *without scroll* condition (*M* = 44.62, *SD* = 6.87); *t* (52) = -3.12, *p* = 0.003, *d*= -0.86. A significant difference in the perceived e-learner satisfaction scores for *with scroll* condition (*M* = 43.66, *SD* = 8.55) and *without scroll* condition (*M* = 50.44, *SD* = 9.62); *t* (52) = -2.73, *p* = 0.009, *d*= -0.75. No significant differences across groups were observed in American students in the dependent measures across the two CBA

Table 4.7

Summary of Results

Dependent Measures	Results
Subjective Cognitive Load	1. Indian students report significantly higher 'cognitive load' during with scroll condition in comparison to without scroll condition. 2. Indian students report significant low 'cognitive load' during assessment in without scroll CBA condition comparison to their American counterparts.
Perceived System Usability	1. Indian students rate 'system usability' of with scroll CBA condition significantly lower in comparison to their American counter parts. 2. Indian students rate 'system quality' of with scroll CBA condition significantly lower in comparison to the without scroll condition.
Perceived Ease of Use	1. Indian students rate 'perceived ease of use' of with scroll CBA condition significantly lower in comparison to their American counter parts. 2. Indian students report significantly higher 'ease of use' during without scroll condition in comparison to with scroll condition.
Perceived System Quality	Indian students rate 'system quality' of with scroll CBA condition significantly lower in comparison to their American counterparts.
Perceived E-learner's Satisfaction	Indian students report significantly higher 'satisfaction' during without scroll condition in comparison to with scroll condition.
Time Taken	1. Indian students take significantly longer period of time to complete the test in with scroll test condition in comparison to their American counterparts. 2. Indian students take significantly longer period of time to complete the test in without scroll CBA condition in comparison to their American counterparts.

test conditions. A Cartesian coordinate graph was plotted by calculating Instructional efficiency of *with scroll* and *without scroll* condition between American students and Indian students. Figure 4.1 depicts the Cartesian coordinate plot of Instructional efficiencies of the test conditions across the two student communities. Figure 4.2 – 4.7 depicts the estimated marginal means of NASA TLX, SUS, PEOU, SQ, PES and TK which has been plotted for with scroll and without scroll conditions across the two student communities.

Discussion: Experiment 1

The results of the multivariate analysis conducted for the *Experiment 1* (With scroll vs. without scroll condition) highlights important facts regarding the effect of the *scrolling* and *without scrolling* features in CBA system. A significant MANOVA interaction effect was

observed. Though significant main effects of interface features and cultural backgrounds were also observed but the significant interaction effect suggests that one level of our criterion variable that is the *interface feature* depends on the level of the other criterion variable i.e., *cultural backgrounds* to effect group differences. Therefore, the hypotheses $\mathbf{H_{1a}}$ and $\mathbf{H_{1b}}$ do not hold importance even though we found significant effects of interface feature and cultural backgrounds alone on the dependent measure. Table 4.6 highlights the hypotheses test results for the interaction effect. The univariate results supports the hypotheses for subjective cognitive load, perceived system usability, perceived ease of use, perceived system quality, perceived e-learner satisfaction and total time taken to complete the test. This indicates that across these dependent measures variability can be explained by the interacting effect of *interface feature* and *cultural background*.

The direction of the interacting effects was investigated by employing independent sample *t*-test on the dependent measures. Table 4.7 depicts the summary of the results that signifies the direction of the effects on the dependent measures across the test conditions. The results of the analysis highlight that Indian students prefer *without scrolling* CBA system over *with scrolling* CBA system. They report significantly high cognitive load for *with scroll* CBA condition. They rated the system usability, perceived ease of use higher of *without scroll* CBA condition in comparison to the *with scroll* CBA condition. They also report higher satisfaction during *without scroll* CBA condition than *with scroll* CBA condition. American students were found to be quicker in both the test conditions in comparison to their Indian counterparts. American students reported lower cognitive load in *with scroll* CBA condition in comparison to their Indian counterparts. Similarly American students perceived system usability, ease of use of *with scroll* CBA condition higher in comparison to their Indian counterparts.

It is important to state that though the American students did not differ significantly in terms of group differences (between the CBA test conditions), but when the significance value is observed from a one tailed significant test perspective (**Appendix- 3A**), different results emerge. It is observed that there is a significant difference in the scores of perceived system usability between *with scroll* and *without scroll* condition. While *with scroll* CBA condition was rated with higher system usability than *without scroll* condition. Likewise, American students took significantly lesser time in completing the test during with scroll CBA condition than *without scroll* CBA condition. Another indication of the preference of American students for the *with scroll* CBA condition can be observed from the instructional efficiency plot (Figure 4.1). Though, there was no significant difference in the subjective cognitive load score across the two test condition- *with* and *without scroll* for American students but when performance (total marks) and corresponding cognitive load are combined together to calculate the instructional

efficiency of the test conditions, the *with scroll* CBA condition was found to have higher instructional efficiency than the *without scroll* condition. Combining the one tailed significant results and the instructional efficiency plotting of the American students (those students who are experts); it can be argued that they prefer CBA systems *with scroll* feature. For the Indian students (those students who are experts), the significant results and the instructional efficiency plotting highlight that Indian students prefer CBA system *without scroll* feature.

4.2.2 Experiment 2: Vertical menu based vs. Step based navigation

The working hypotheses which were tested in the experiment vertical menu based navigation vs. step menu based navigation are listed below:

H_{2a}: Navigation alone would affect performance of students, perceived cognitive load, perceived system usability, perceived ease of use, perceived usefulness, system quality, e-learner's satisfaction, willingness to continue the usage and relative advantage in a summative CBA environment.

H_{2b}: Cultural background alone would affect performance of students, perceived cognitive load, perceived system usability, perceived ease of use, perceived usefulness, system quality, e-learner's satisfaction, willingness to continue the usage and relative advantage in a summative CBA environment.

H_{2c}: The interacting effect of cultural background and navigation alone would affect performance of students, perceived cognitive load, perceived system usability, perceived ease of use, perceived usefulness, system quality, e-learner's satisfaction, willingness to continue the usage and relative advantage in a summative CBA environment.

Table 4.8 and table 4.9 highlight the descriptive statistics of the participants and the dependent measures that were captured during the experiment. In the vertical menu based CBA condition 27 students (15 male and 12 female; M_{Age} = 23.0, SD_{Age} = 2.68; M_{Marks} = 69.62, SD_{Marks} = 8.98) from India participated, while 27 students (11 male and 16 female; M_{Age} = 18.48, SD_{Age} = 0.89; M_{Marks} = 72.37, SD_{Marks} = 9.51) from USA participated. For the step based navigation CBA condition 26 students (16 male and 10 female; M_{Age} = 22.65, SD_{Age} = 2.33; M_{Marks} = 71.34, SD_{Marks} = 8.41) from India participated, while 29 students (10 male and 19 female; M_{Age} = 18.72, SD_{Age} = 1.62; M_{Marks} = 71.03, SD_{Marks} = 8.58) from USA participated. All the four groups were homogeneous in terms of their age (between interface feature and not culture) and marks obtained in the last qualifying exam (See Appendix 4, page 240).

Table 4.8

Distribution of Gender, age and marks obtained in the last qualifying exam across Exam conditions and the countries.

	Vertical menu based navigation		Step navigation	
	India	**USA**	**India**	**USA**
Male	15	11	16	10
Female	12	16	10	19
Total	27	27	26	29
Age	23.00 (*M*)	18.48(*M*)	22.65(*M*)	18.72(*M*)
	2.68 (*SD*)	0.89(*SD*)	2.33(*SD*)	1.62(*SD*)
Marks obtained in last qualifying exam	69.62(*M*)	72.37(*M*)	71.34(*M*)	71.03(*M*)
	8.98(*SD*)	9.51(*SD*)	8.41(*SD*)	8.58(*SD*)

Table 4.9

Descriptive statistics of the dependent variables.

	Vertical menu based navigation				Step navigation			
	India		USA		India		USA	
	M	***S.D***	***M***	***S.D***	***M***	***S.D***	***M***	***S.D***
NASATLX	31.89	18.63	27.55	13.23	33.05	11.97	26.00	13.01
SUS	72.48	15.92	82.50	12.88	74.61	14.02	85.43	10.98
PEOU	17.11	2.89	17.78	3.25	16.88	3.30	16.45	3.99
PEU	12.04	2.38	12.89	2.81	11.62	2.93	12.10	3.02
SQ	46.22	6.10	44.30	7.37	44.92	5.23	45.21	7.11
RA	7.00	1.88	6.30	1.96	7.12	1.90	6.14	1.73
PES	49.70	8.27	47.93	11.31	49.73	11.04	46.69	12.68
WCU	8.07	1.49	6.96	2.05	7.92	1.72	7.03	2.23
TM	6.00	1.71	5.93	1.59	5.62	1.68	5.90	1.21
TK	11.80	3.54	6.61	2.40	6.29	3.83	5.58	1.71

Table 4.10

Summarized results of the tests of Normality (Q-Q plots).

Dependent Measures	Vertical menu		Step based	
	India	**USA**	**India**	**USA**
NASA TLX	Slight deviations	Perfectly Normal	Almost Normal	Perfectly Normal
SUS	Almost Normal	Almost Normal	Almost Normal	Almost Normal
PEOU	Slight deviations	Slight deviations	Slight deviations	Slight deviations
PEU	Almost Normal	Almost Normal	Almost Normal	Slight deviations
SQ	Almost Normal	Slight deviations	Almost Normal	Almost Normal
RA	Perfectly Normal	Perfectly Normal	Perfectly Normal	Perfectly Normal
WCU	Perfectly Normal	Almost Normal	Slight deviations	Almost Normal
PES	Perfectly Normal	Almost Normal	Almost Normal	Slight deviations
TM	Perfectly Normal	Perfectly Normal	Perfectly Normal	Perfectly Normal
TK	Almost Normal	Slight deviations	Almost Normal	Almost Normal

The inferential statistical analysis was carried out first starting with the tests of Normality. Q-Q plots were generated to test the normality of the data distribution across the groups for all the dependent measures. **Appendix-3B** enlists the Q-Q plots of the data distributions. The observations are summarized in table 4.10. Thereafter, the assumption of homogeneity of variance-covariances were tested using Box's Test. Box's test of equality of covariance matrices highlights that it is significant at $p<0.005$. Therefore, the assumption of homogeneity of variance-covariance matrices was violated. The Q-Q plots revealed that there are slight deviations from Normality for some of the dependent measures and the Box's test revealed that the assumption of the homogeneity of variance-covariance matrices was violated. Therefore, the Pillai's statistic was used to derive the conclusions as it is robust to the violations of these assumptions (Haase & Ellis, 1987).

The 2x2 factorial multivariate tests (MANOVA) conducted on the data sets (Table 4.11) revealed that there was a statistically significant main effect for interface feature, F (10, 96) = 4.00, $p < 0.001$; Pillai's Trace = 0.29; partial η^2 = 0.29. There was also a statistically significant main effect for Country, F (10, 96) = 7.49, $p < 0.001$; Pillai's Trace = 0.44; partial η^2 = 0.44. There was marginally significant interaction effect between Country and interface feature. Univariate analysis conducted (Table 4.12) reveals that the group membership, interface feature (navigational mechanism) had a significant effect on time taken, F (1, 105) = 33.05, $p <$ 0.001; partial η^2 =0.23. Univariate analysis conducted for the group membership, cultural backgrounds had a significant effect on subjective cognitive load, F (1, 105) =4.24, $p < 0.05$; partial η^2 = 0.03; perceived system usability F (1, 105) =16.16, $p < 0.001$; partial η^2 =0.13;

Table 4.11

Multivariate statistics

Effect		Value	*F*	*Hypothesis df*	*Error df*	Sig.	Partial Eta Squared	Observed Power[c]
Exam	**Pillai's**	0.29	4.00	10.00	96.00	0.00	0.29	1.00
Country	**Pillai's**	0.44	7.49	10.00	96.00	0.00	0.44	1.00
Exam * Country	**Pillai's**	0.16	1.81	10.00	96.00	0.07	0.16	0.81

Table 4.12

Univariate statistics of the dependent measures on the Independent measures

Source	Dependent Variable	Type III Sum of Squares	*df*	Mean Square	*F*	Sig.	Partial Eta Squared
Exam	**NASATLX**	1.03	1	1.03	0.00	0.94	0.00
	SUS	174.36	1	174.36	0.95	0.33	0.00
	PEOU	16.47	1	16.47	1.43	0.23	0.01
	PEU	9.91	1	9.91	1.27	0.26	0.01
	SQ	1.03	1	1.03	0.02	0.88	0.00
	RA	0.01	1	0.01	0.00	0.95	0.00
	PES	9.95	1	9.95	0.08	0.77	0.00
	WCU	0.04	1	0.04	0.01	0.91	0.00
	TM	1.17	1	1.17	0.48	0.49	0.00
	TK	290.90	1	290.90	33.05	0.00	0.23
Country	**NASATLX**	883.20	1	883.20	4.24	0.04	0.03
	SUS	2952.98	1	2952.98	16.16	0.00	0.13
	PEOU	0.36	1	0.36	0.03	0.86	0.00
	PEU	12.21	1	12.21	1.56	0.21	0.01
	SQ	18.34	1	18.34	0.43	0.51	0.00
	RA	19.22	1	19.22	5.52	0.02	0.05
	PES	157.95	1	157.95	1.31	0.25	0.01
	WCU	27.20	1	27.20	7.52	0.00	0.06
	TM	0.29	1	0.29	0.12	0.72	0.00
	TK	236.91	1	236.91	26.91	0.00	0.20

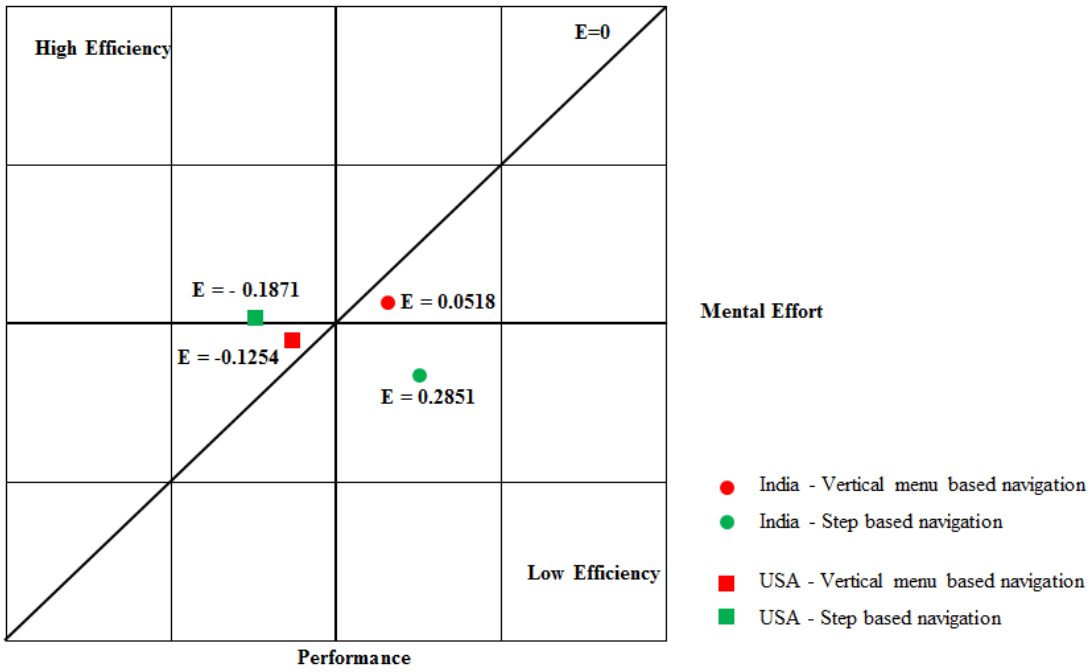


Figure 4.8. Instructional efficiency of vertical menu based navigation and step based navigation CBA conditions between Indian and American students.

relative advantage $F(1, 105) = 5.52$, $p < 0.05$; partial $\eta^2 = 0.05$; willingness to continue to use $F(1, 105) = 7.52$, $p < 0.001$; partial $\eta^2 = 0.06$; $F(1, 105) = 26.91$, $p < 0.001$; partial $\eta^2 = 20$.

To identify the direction of the main effects pair-wise comparisons have been carried out (see **Appendix-3B**). The pair wise comparisons highlight that time taken in *vertical menu based navigation* was significantly higher than *step based navigation*. Subjective cognitive load reported was significantly higher for Indian students than their American counterparts. Similarly, relative advantage was reported significantly higher by Indians than their American counterparts. Perceived system usability was significantly lower for Indian students than their American counterparts. Time taken to complete the test was significantly lower by American students than their Indian counter parts. Figure 4.8 highlights the Cartesian coordinate plot of Instructional efficiency across the two interface features for the Indian and the American Students. It is highlighted that both the CBA systems (*vertical menu based navigation and step based navigation*) lie in the high efficiency zone for the American students but for the Indian students they lay in the low efficiency zone. Figure 4.9 - 4.13 depicts the marginal means of the dependent measures NASA TLX, TK, SUS, RA, and WCU.

Discussion: Experiment 2

The results of the multivariate analysis conducted for the Experiment 2 (Vertical menu based navigation vs. step based navigation) highlights important facts regarding the effect of the vertical menu based navigational mechanism and step based navigational mechanism features in

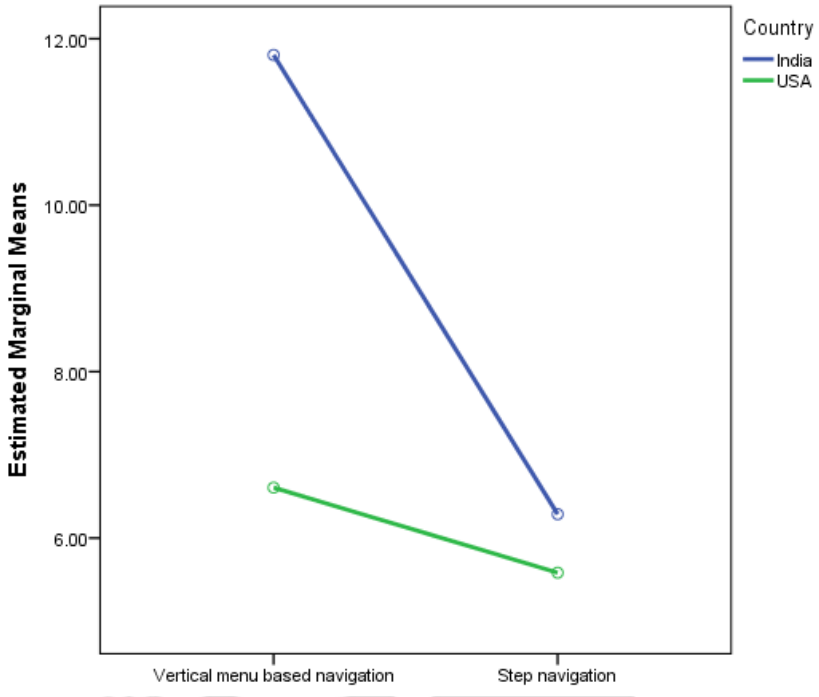


Figure 4.9. Estimated Marginal means of Time Taken across the test conditions and countries.

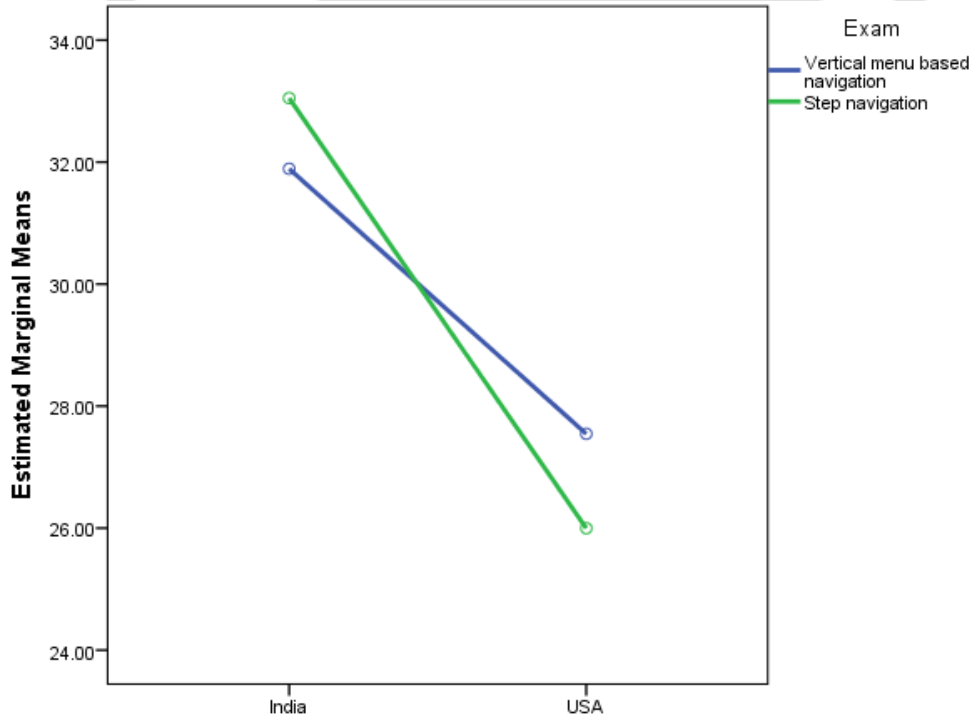


Figure 4.10. Estimated Marginal means of NASA TLX across the test conditions and countries.

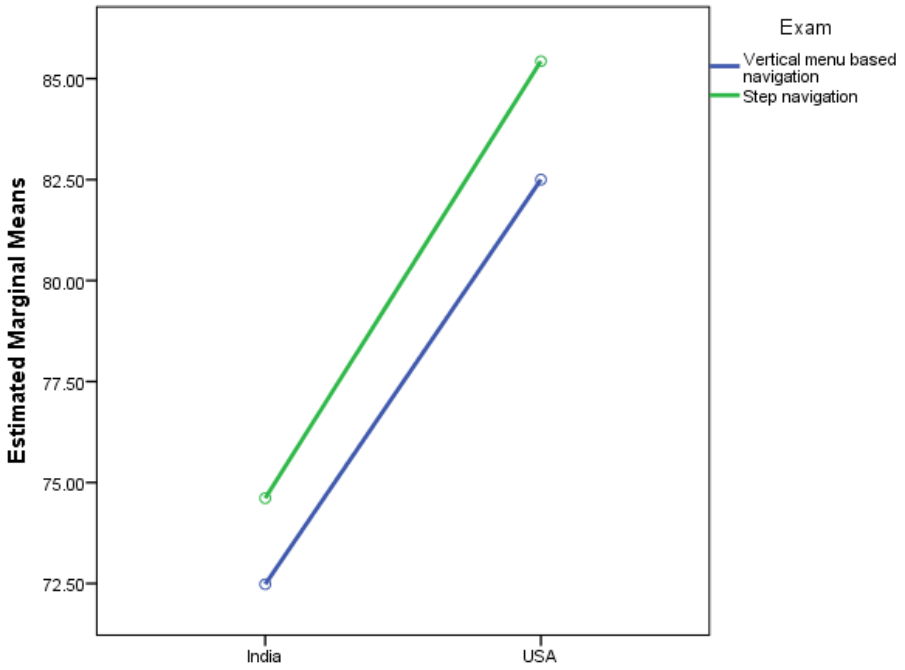


Figure 4.11. Estimated Marginal means of SUS across the test conditions and countries.

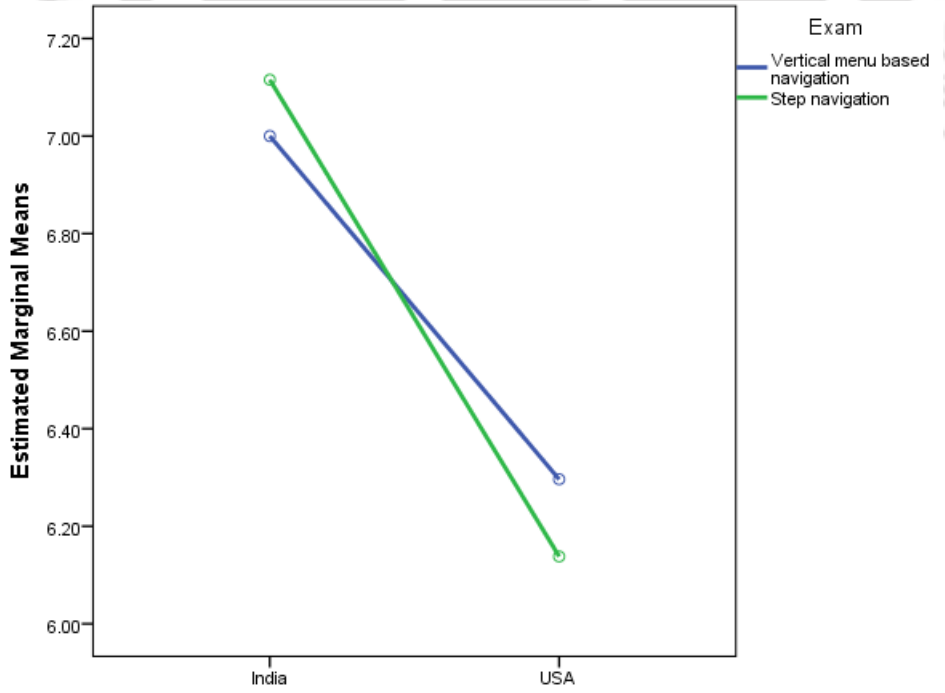


Figure 4.12. Estimated Marginal means of RA across the test conditions and countries.

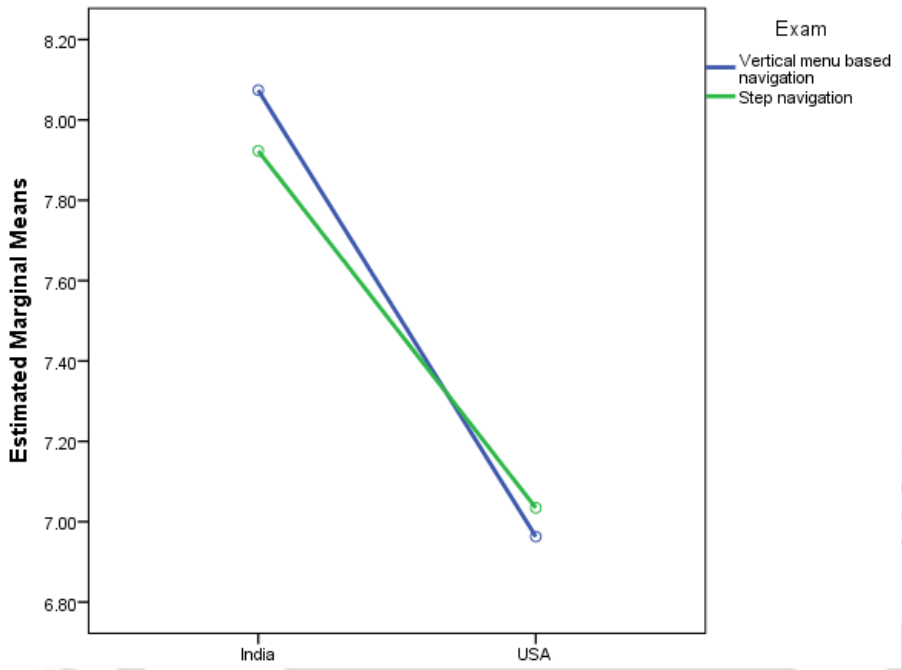


Figure 4.13. Estimated Marginal means of WCU across the test conditions and countries.

a CBA system. A significant MANOVA main effect for navigational mechanisms and cultural background was observed. Univariate analysis conducted to assess the effect of the criterion variables on the dependent measures reveals that navigational mechanisms affect the amount of time taken by the students among the two CBA conditions. A pair wise analysis revealed the direction of the effect.

Students took longer time in completing the CBA with vertical menu based navigational mechanisms. For the significant main effect of cultural backgrounds observed in the multivariate analysis, series of univariate analysis on the dependent measures revealed that Cultural background affects subjective cognitive load, perceived system usability, relative advantage, willingness to continue to use and time taken to complete the test. Table 4.13 highlights the hypothesis testing results and table 4.14 enlists the summary of the findings.

While Indian students had reported subjective cognitive load, rate perceived system usability, relative advantage, willingness to continue to use higher as compared to their American counterparts. American students have been found to complete both the formats of the CBA at a much faster rate than their Indian counterparts.

Table 4.13

Hypothesis test results

Hypothesis $\mathbf{H_{2a}}$ = Interface feature affects	Supported
Subjective Cognitive Load	Not Supported
Perceived System Usability	Not Supported
Perceived Ease of Use	Not Supported
Perceived Usefulness	Not supported
Perceived System Quality	Not Supported
Relative Advantage	Not Supported
Perceived E-learner's Satisfaction	Not Supported
Willingness to Continue to Use	Not Supported
Total Marks	Not Supported
Time Taken	Supported
Hypothesis $\mathbf{H_{2b}}$ = Cultural background affects	Supported
Subjective Cognitive Load	Supported
Perceived System Usability	Supported
Perceived Ease of Use	Not Supported
Perceived Usefulness	Not supported
Perceived System Quality	Not Supported
Relative Advantage	Supported
Perceived E-learner's Satisfaction	Not Supported
Willingness to Continue to Use	Supported
Total Marks	Not Supported
Time Taken	Supported
Hypothesis $\mathbf{H_{2c}}$ = Cultural background X Interface feature affects	Not Supported

Overall, the investigation found that the students (experts) of the two countries differ in terms of their individual characteristics when provided with two formats of CBA conditions. The instructional efficiency plotting suggests that for American students both the CBA conditions fall in high instructional efficiency zone while for their Indian counterparts both fall in the low instructional efficiency zone. So, for both the student groups navigational mechanisms do not affect their performance and their subjective evaluation of the CBA systems.

Table 4.14

Summary of Results

Dependent Measures	Results
Subjective Cognitive Load	Overall Indian students report significantly higher 'cognitive load' for both the CBA conditions in comparison to their American counterparts.
Perceived System Usability	Overall Indian students report significantly lower 'system usability' for both the CBA conditions in comparison to their American counterparts.
Relative Advantage	Overall Indian students report significantly higher 'relative advantage' for both the CBA conditions in comparison to their American counterparts.
Willingness to continue to use	Overall Indian students report significantly higher 'willingness to continue to use' for both the CBA conditions in comparison to their American counterparts.
Time Taken	Time taken by students to complete the CBA with vertical menu based interface feature is significantly higher than in CBA with step based navigational feature. Overall Indian students take significantly longer time to complete the test in comparison to their American counterparts.

4.2.3 Experiment 3: Question Customization vs. Time Customization

The working hypotheses which were tested in this experiment question customization vs. time customization are as follows:

H_{3a}: Customization feature alone would affect performance of students, perceived cognitive load, perceived system usability, perceived ease of use, perceived usefulness, system quality, e-learner's satisfaction, willingness to continue the usage and relative advantage in a summative CBA environment.

H_{3b}: Cultural background alone would affect performance of students, perceived cognitive load, perceived system usability, perceived ease of use, perceived usefulness, system quality, e-learner's satisfaction, willingness to continue the usage and relative advantage in a summative CBA environment.

H_{3c}: The interacting effect of cultural background and customization feature alone would affect performance of students, perceived cognitive load, perceived system usability,

Table 4.15

Distribution of Gender, age and marks obtained in the last qualifying exam across Exam conditions and the countries.

	Question Customization		Time Customization	
	India	**USA**	**India**	**USA**
Male	17	16	18	14
Female	12	18	11	20
n	29	34	29	34
Age	22.51 (*M*) 2.35 (*SD*)	19.44(*M*) 0.70(*SD*)	23.10(*M*) 2.22(*SD*)	18.55(*M*) 0.89(*SD*)
Marks obtained in last qualifying exam	70.12(*M*) 5.90(*SD*)	65.30(*M*) 8.55(*SD*)	71.85(*M*) 7.45(*SD*)	67.15(*M*) 6.55(*SD*)

Table 4.16

Descriptive statistics of the dependent measures.

	Question Customization				Time Customization			
	India		USA		India		USA	
	M	***S.D***	***M***	***S.D***	***M***	***S.D***	***M***	***S.D***
NASATLX	34.78	13.65	34.42	12.31	39.96	20.50	46.78	16.12
SUS	72.07	16.24	81.47	13.43	63.14	19.59	69.26	16.62
PEOU	16.10	2.77	16.88	3.34	14.66	3.47	13.97	5.12
PEU	11.10	2.76	12.24	3.13	9.83	2.88	9.94	4.18
SQ	40.31	7.19	45.85	5.68	39.07	8.67	40.18	7.91
RA	6.55	1.68	5.94	1.84	6.52	1.82	4.82	1.93
PES	46.17	9.11	44.97	11.73	43.17	10.99	40.74	12.38
WCU	7.24	1.62	7.03	1.93	6.53	2.39	6.21	2.27
TM	5.21	1.90	6.29	1.57	4.90	2.43	4.74	2.19
TK	8.00	3.76	7.03	3.05	11.58	4.11	8.97	3.77

Table 4.17

Summarized results of the tests of Normality (Q-Q plots).

Dependent Measures	Question customization		Time customization	
	India	**USA**	**India**	**USA**
NASA TLX	Almost Normal	Almost Normal	Almost Normal	Almost Normal
SUS	Almost Normal	Slight deviations	Almost Normal	Almost Normal
PEOU	Slight deviations	Slight deviations	Almost Normal	Almost Normal
PEU	Almost Normal	Slight deviations	Almost Normal	Almost Normal
SQ	Almost Normal	Almost Normal	Almost Normal	Almost Normal
RA	Slight deviations	Almost Normal	Almost Normal	Slight deviations
WCU	Slight deviations	Slight deviations	Slight deviations	Slight deviations
PES	Slight deviations	Slight deviations	Almost Normal	Almost Normal
TM	Almost Normal	Almost Normal	Almost Normal	Almost Normal
TK	Slight deviations	Almost Normal	Slight deviations	Almost Normal

perceived ease of use, perceived usefulness, system quality, e-learner's satisfaction, willingness to continue the usage and relative advantage in a summative CBA environment.

Table 4.15 and table 4.16 highlight the descriptive statistics of the participants and the dependent measures that were captured during the experiment. In the question customization CBA condition 29 students (17 male and 12 female; M_{Age} = 22.51, SD_{Age} = 2.35; M_{Marks} = 70.12, SD_{Marks} = 5.90) from India participated, while 34 students (16 male and 18 female; M_{Age} = 19.44, SD_{Age} = 0.70; M_{Marks} = 65.30, SD_{Marks} = 8.55) from USA participated. For the time customization CBA condition 29 students (18 male and 11 female; M_{Age} = 23.10, SD_{Age} = 2.22; M_{Marks} = 71.85, SD_{Marks} = 7.45) from India participated, while 34 students (14 male and 20 female; M_{Age} = 18.55, SD_{Age} = 0.89; M_{Marks} = 67.15, SD_{Marks} = 6.55) from USA participated. All the four groups were homogeneous in terms of their age (between interface feature and not between culture) and marks obtained in the last qualifying exam (see Appendix 4, page 240).

The inferential statistical analysis was carried out starting first with the tests of Normality. Q-Q plots were generated to test the normality of the data distribution across the groups for all the dependent measures. **Appendix-3C** enlists the Q-Q plots of the data distributions. The observations are summarized in table 4.17. Thereafter, the assumption of homogeneity of variance-covariances were tested using Box's Test. Box's test of equality of covariance matrices highlights that it is significant at p<0.005. Therefore, the assumption of homogeneity of variance-covariance matrices was violated. The Q-Q plots revealed that there are slight deviations from Normality for some of the dependent measures and the Box's test revealed that the assumption of the homogeneity of variance-covariance matrices was violated.

Table 4.18

Multivariate statistics

Effect		Value	*F*	Hypothesis df	Error df	Sig.	Partial Eta Squared	Observed Power
Exam	**Pillai's Trace**	0.25	3.78	10	113	0.00	0.25	0.99
Country	**Pillai's Trace**	0.33	5.60	10	113	0.00	0.33	1.00
Exam * Country	**Pillai's Trace**	0.09	1.06	10	113	0.40	0.09	0.53

Table 4.19

Univariate statistics of the dependent measures based on the independent factors.

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	*F*	Sig.	Partial Eta Squared	Observed Power[k]
Exam	**NASATLX**	2405.51	1	2405.51	9.60	0.00	0.07	0.87
	SUS	3496.15	1	3496.15	12.85	0.00	0.10	0.94
	PEOU	148.76	1	148.76	10.17	0.00	0.08	0.89
	PEU	99.73	1	99.73	9.05	0.00	0.07	0.85
	SQ	374.50	1	374.50	6.84	0.01	0.05	0.74
	RA	10.39	1	10.39	3.11	0.08	0.02	0.42
	PES	409.66	1	409.66	3.27	0.07	0.03	0.43
	WCU	18.33	1	18.33	4.25	0.04	0.03	0.53
	TM	27.34	1	27.34	6.60	0.01	0.05	0.72
	TK	238.64	1	238.64	17.69	0.00	0.13	0.99
Country	**NASATLX**	326.78	1	326.78	1.30	0.26	0.01	0.21
	SUS	1886.95	1	1886.95	6.93	0.01	0.05	0.74
	PEOU	0.07	1	0.07	0.00	0.95	0.00	0.05
	PEU	12.14	1	12.14	1.10	0.30	0.01	0.18
	SQ	346.07	1	346.07	6.32	0.01	0.05	0.70
	RA	41.55	1	41.55	12.45	0.00	0.09	0.94
	PES	103.62	1	103.62	0.83	0.37	0.01	0.15
	WCU	2.29	1	2.29	0.53	0.47	0.00	0.11
	TM	6.71	1	6.71	1.62	0.21	0.01	0.24
	TK	100.31	1	100.31	7.44	0.01	0.06	0.77

Therefore, the Pillai's statistic was used to derive the conclusions as it is robust to the violations of these assumptions (Haase & Ellis, 1987).

The 2x2 factorial multivariate tests (MANOVA) conducted on the data sets (Table 4.18) revealed that there was a statistically significant main effect for interface feature, F (10, 113) = 3.78, $p < 0.001$; Pillai's Trace = 0.25; partial η^2 = 0.25. There was also a statistically significant main effect for Country, F (10, 113) = 5.60, $p < 0.001$; Pillai's Trace = 0.33; partial η^2 = 0.33. There was no statistically significant interaction effect between Country and interface feature.

Univariate analysis conducted reveals that the group membership (Table 4.19), interface feature (customization) had a significant effect on subjective cognitive load, F (1, 122) =9.60, $p < 0.001$; partial η^2 =0.07; perceived system usability F (1, 122) =12.85, $p < 0.001$; partial η^2 =0.10; perceived ease of use F (1, 122) = 10.17, $p < 0.001$; partial η^2 = 0.08; perceived usefulness F (1, 122) = 9.05, $p < 0.001$; partial η^2 = 0.07; perceived system quality F (1, 122) = 6.84, $p < 0.05$; partial η^2 = 0.05; willingness to continue to use F (1, 122) = 4.25, $p < 0.05$; partial η^2 = 0.03; total marks F (1, 122) = 6.60, $p < 0.05$; partial η^2 = 0.05; time taken F (1, 122) = 17.69, $p < 0.001$; partial η^2 = 0.13. Univariate analysis conducted for the group membership, cultural backgrounds had a significant effect on perceived system usability, F (1, 122) = 6.93, $p < 0.05$; partial η^2 = 0.05; perceived system quality F (1, 122) = 6.32, $p < 0.05$; partial η^2 = 0.05; relative advantage F (1, 122) = 12.45, $p < 0.001$; partial η^2 = 0.09; time taken F (1, 122) = 7.44, $p < 0.05$; partial η^2 = 0.06.

To identify the direction of the main effects pair-wise comparisons have been carried out (see **Appendix-3C**). The pair wise comparisons for the customization effects highlight that subjective cognitive load for time customization was significantly higher in comparison to the question customization CBA condition. The question customization CBA condition has significant higher ratings for perceived system usability, perceived ease of use, perceived usefulness, perceived system quality, willingness to continue to use and total marks against the time customization CBA condition. Students were significantly was faster completing the CBA session with question customization than with time customization CBA condition.

The pairwise comparisons for the main effect of cultural backgrounds on perceived system usability, perceived system quality, relative advantage and time taken highlighted that the perceived system usability score and the perceived system quality score for the Indians were significantly lower than their American counterparts. Indian

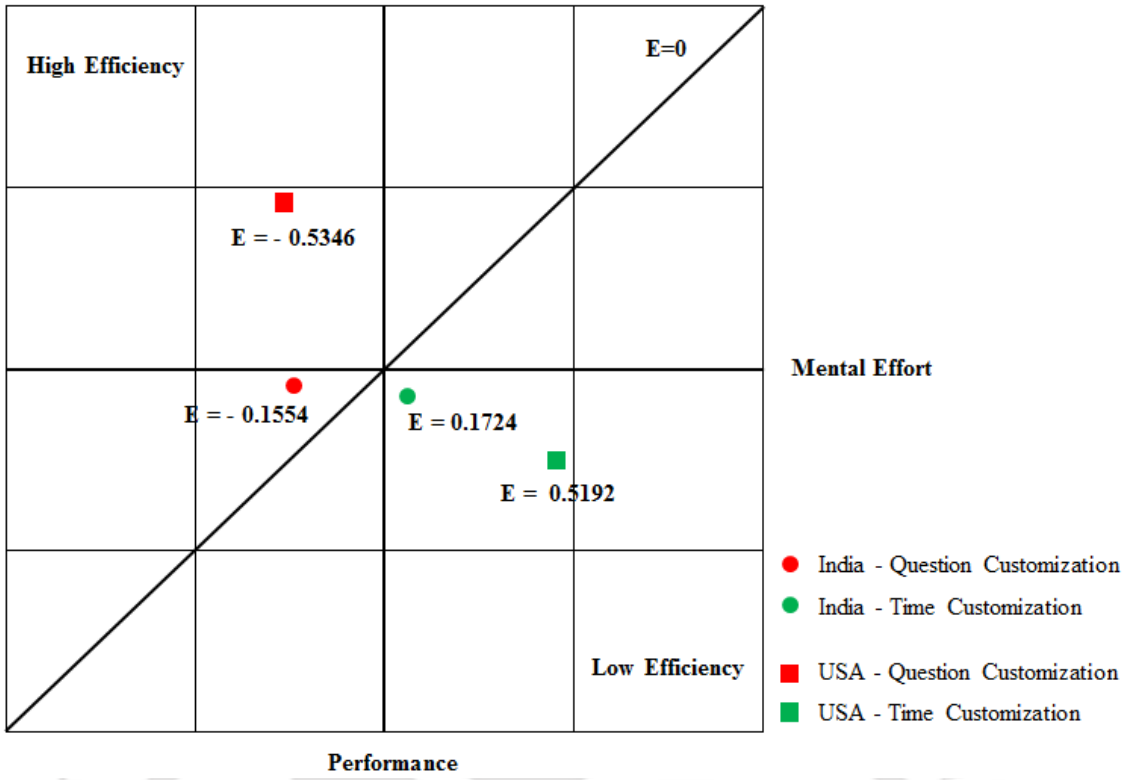


Figure 4.14. Instructional efficiency of Question customization and Time customization CBA conditions between Indian and American students.

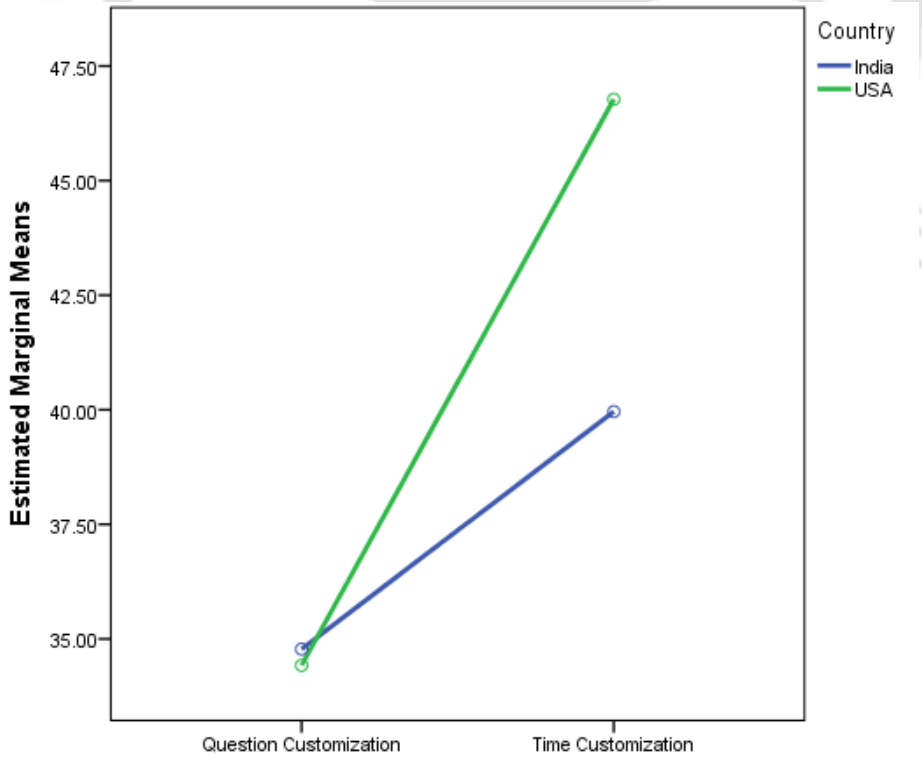


Figure 4.15. Estimated Marginal means of NASA TLX across the test conditions and countries.

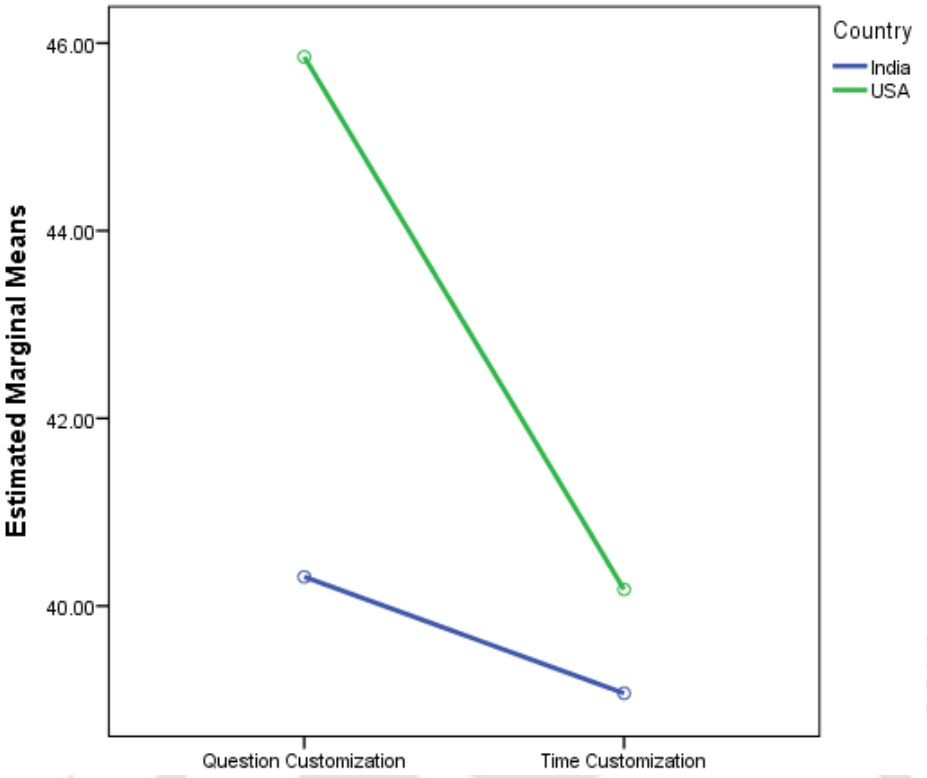


Figure 4.16. Estimated Marginal means of SQ across the test conditions and countries.

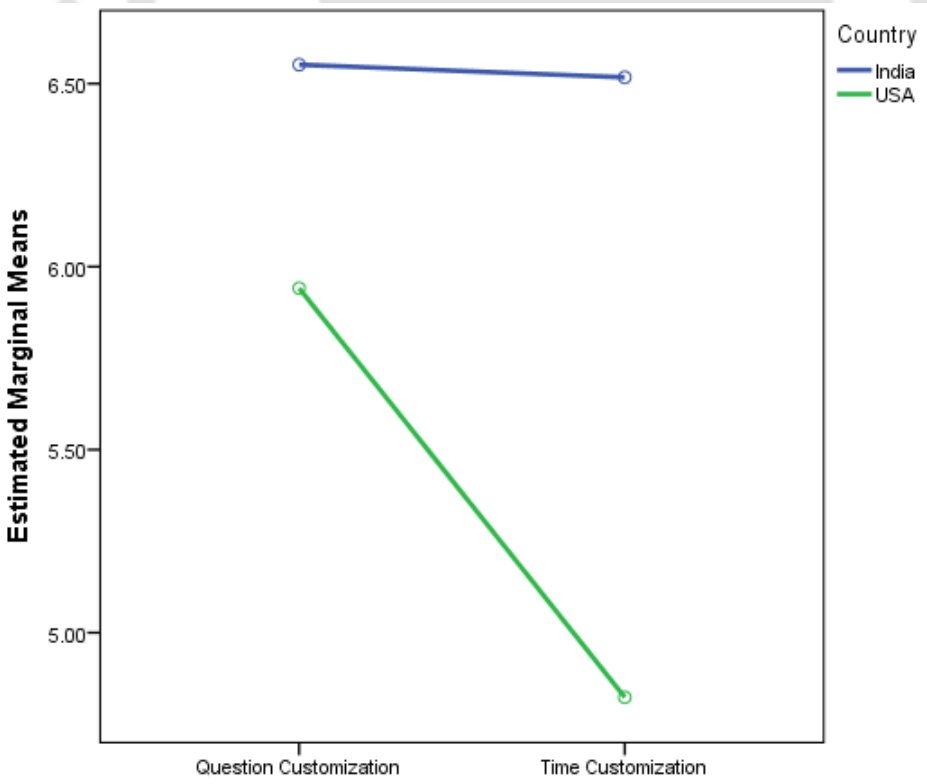


Figure 4.17. Estimated Marginal means of RA across the test conditions and countries.

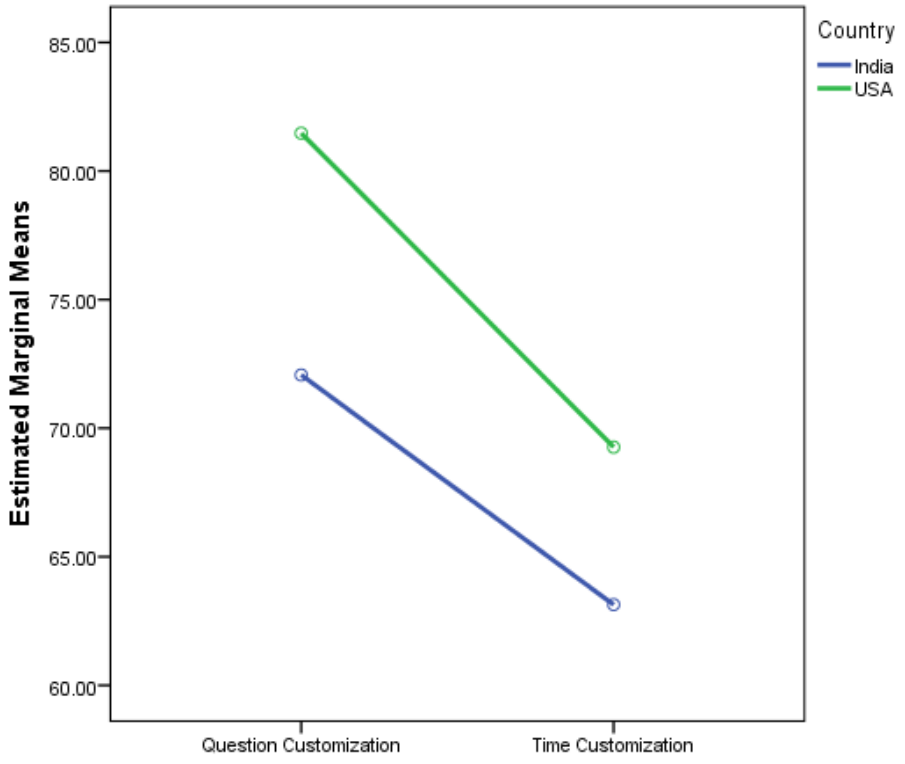


Figure 4.18. Estimated Marginal means of SUS across the test conditions and countries.

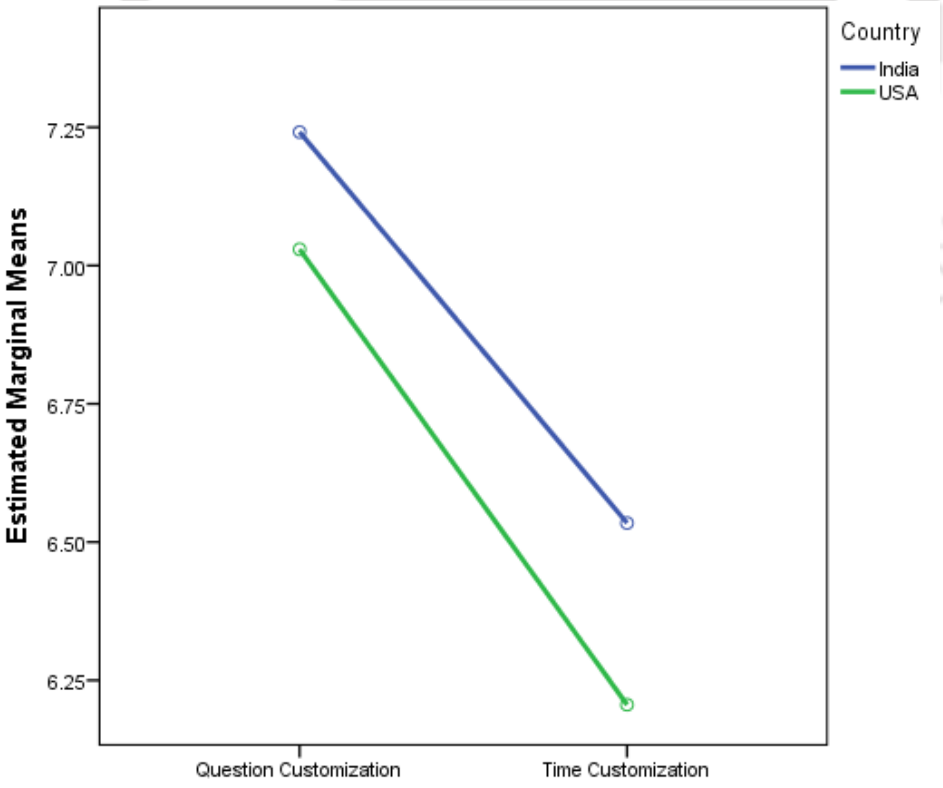


Figure 4.19. Estimated Marginal means of WCU across the test conditions and countries.

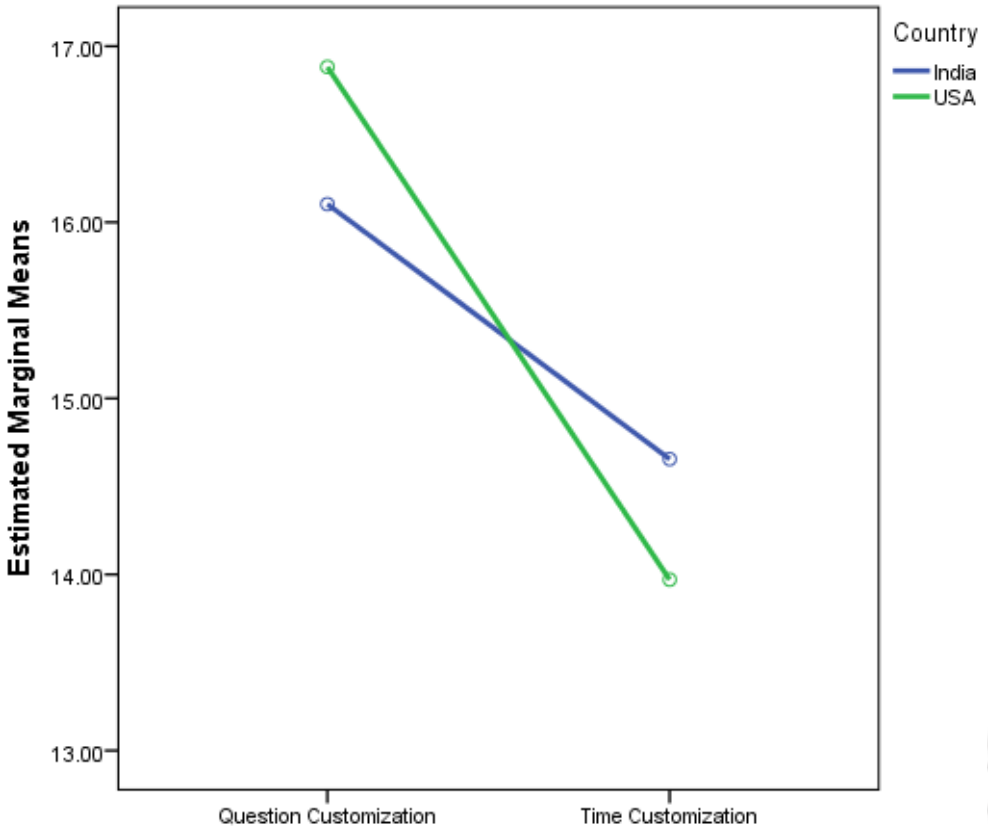


Figure 4.20. Estimated Marginal means of PEOU across the test conditions and countries.

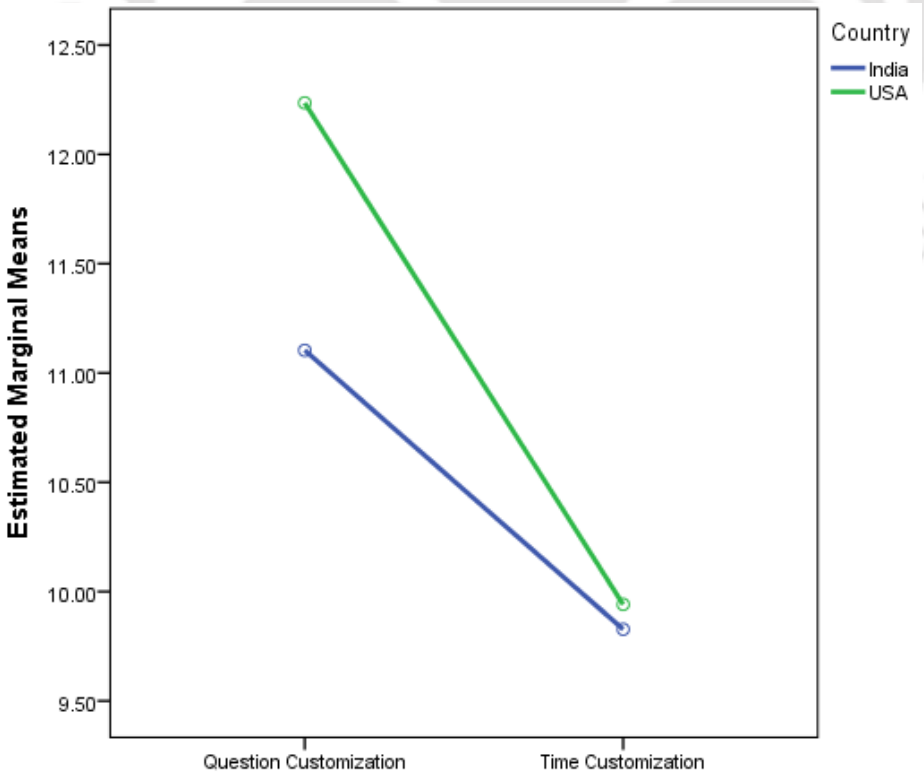


Figure 4.21. Estimated Marginal means of PEU across the test conditions and countries.

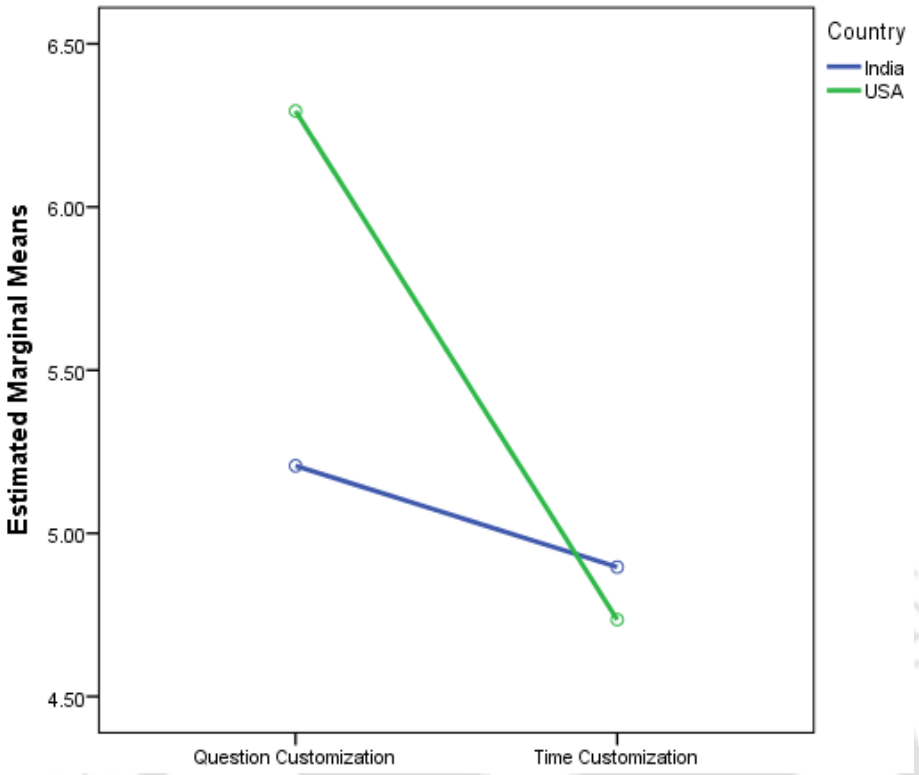


Figure 4.22. Estimated Marginal means of TM across the test conditions and countries.

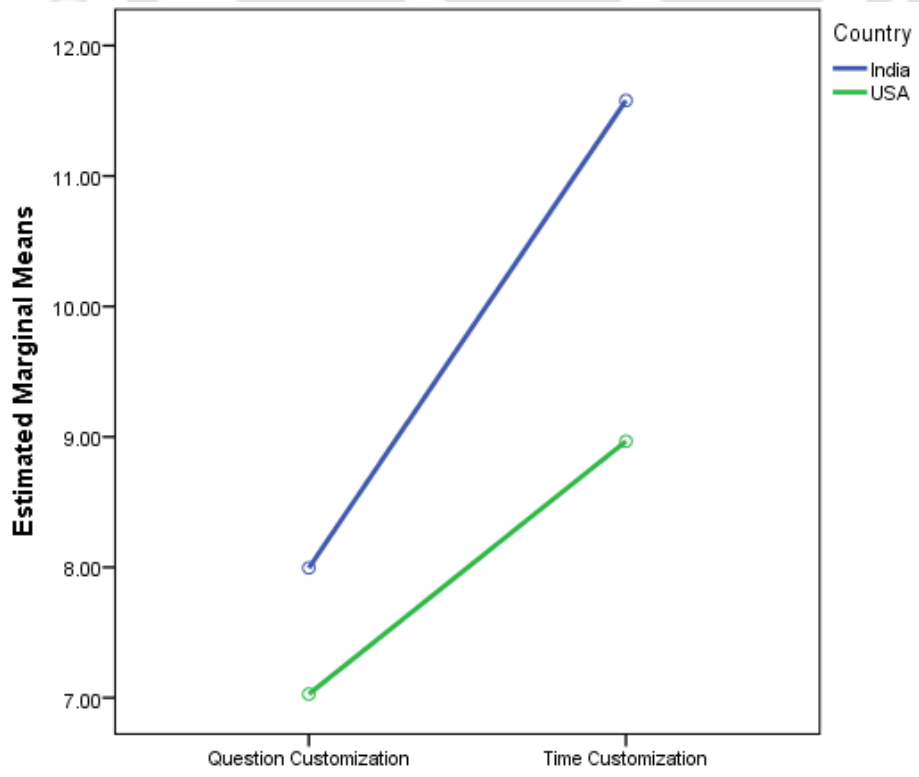


Figure 4.23. Estimated Marginal means of TK across the test conditions and countries.

Table 4.20

Hypotheses testing

Hypothesis $\mathbf{H_{3a}}$ = Interface feature affects	Supported
Subjective Cognitive Load	Supported
Perceived System Usability	Supported
Perceived Ease of Use	Supported
Perceived Usefulness	Supported
Perceived System Quality	Supported
Relative Advantage	Not Supported
Perceived E-learner's Satisfaction	Not Supported
Willingness to Continue to Use	Supported
Total Marks	Supported
Time Taken	Supported
Hypothesis $\mathbf{H_{3b}}$ = Cultural background affects	Supported
Subjective Cognitive Load	Not Supported
Perceived System Usability	Supported
Perceived Ease of Use	Not Supported
Perceived Usefulness	Not supported
Perceived System Quality	Supported
Relative Advantage	Supported
Perceived E-learner's Satisfaction	Not Supported
Willingness to Continue to Use	Not Supported
Total Marks	Not Supported
Time Taken	Supported
Hypothesis $\mathbf{H_{3c}}$ = Cultural background X Interface feature affects	Not Supported

students reported relative advantage of CBA conditions relatively higher than their American counterparts. American students took significantly less time in completing the CBA session as compared to their Indian counter parts. Figure 4.14 highlights the Instructional efficiency of the CBA conditions across the two student communities. It has been found that for both the communities *time customization* interface feature based CBA system has less instructional efficiency compared to *question customization* interface feature based CBA system which has high instructional efficiency. Figure 4.15 -4.23 highlights the marginal means of the dependent measures across the two CBA conditions among Indian and American students.

Table 4.21

Summary of Results

Dependent Measures	Results
Subjective Cognitive Load	Overall 'subjective cognitive load' is significantly higher in time customization CBA condition than question customization CBA condition.
Perceived System Usability	Overall 'perceived system usability' is significantly higher in question customization CBA condition than time customization CBA condition. Overall 'perceived system usability' is significantly lower for Indians than their American counterparts.
Perceived Ease of Use	Overall 'perceived ease of use' is significantly higher in question customization CBA condition than time customization CBA condition.
Perceived Usefulness	Overall 'perceived usefulness' is significantly higher in question customization CBA condition than time customization CBA condition.
Perceived System Quality	Overall 'perceived system quality' is significantly higher in question customization CBA condition than time customization CBA condition. Overall 'perceived system quality' is significantly lower for Indians than their American counterparts.
Relative Advantage	Overall 'relative advantage' is significantly higher for Indians than their American counterparts.
Willingness to continue to use	Overall 'willingness to continue' to use is significantly higher in question customization CBA condition than time customization CBA condition.
Total Marks	Overall 'total marks' are significantly higher in question customization CBA condition than time customization CBA condition.
Time Taken	Overall 'total time taken' to complete question customization based CBA session is significantly lower as compared to time customization based CBA session. 'Time taken' by Indian to complete both the assessment sessions were significantly higher than their American counterparts.

Discussion: Experiment 3

The results of the multivariate analysis conducted for the *Experiment 3* (question customization vs. time customization) highlights important facts regarding the effect of the customization features in a CBA system. A significant MANOVA main effect for customization features (*time and question*) and cultural background was observed. Univariate analysis conducted to assess the effect of the criterion variables on the dependent measures reveals that customization features affect subjective cognitive load, perceived system usability, perceived ease of use, perceived usefulness, perceived system quality, willingness to continue to use, total

marks and time taken to complete the assessment session. Table 4.20 highlights the hypotheses testing results and table 4.21 enlists the summary of the findings.

A pair wise analysis revealed the direction of the effect. Students reported higher cognitive load for the CBA session with *time customization* feature, they rated the system usability, ease of use, usefulness, and system quality for the CBA session with question customization feature higher compared to the CBA session with *time customization* feature. It was observed that students with *question customization* based CBA session scored higher than the *time customization* based CBA session. This supports our claim that interface features can affect students' performance. Students from both the countries (India and USA) struggled with the *time customization* feature based CBA session. And as a result of this their performance were affected. This effect of the interface feature can be labelled as *"presentation format effect"*. The two student communities (Indian and American) who participated in the experiments differed in terms of their ratings for the dependent measures like perceived system usability, perceived system quality and relative advantage. While American students took less time to complete the assessment in both the formats their Indian counterparts took significant longer time to complete the assessment session for both the formats.

4.2.4 Experiment 4: Single question without customization vs. Question Customization

The working hypotheses which were tested in this experiment Single question without customization vs. Question Customization are as follows:

H_{4a}: Customization alone would affect performance of students, perceived cognitive load, perceived system usability, perceived ease of use, perceived usefulness, system quality, e-learner's satisfaction, willingness to continue the usage and relative advantage in a summative CBA environment.

H_{4b}: Cultural background alone would affect performance of students, perceived cognitive load, perceived system usability, perceived ease of use, perceived usefulness, system quality, e-learner's satisfaction, willingness to continue the usage and relative advantage in a summative CBA environment.

H_{4c}: The interacting effect of cultural background and customization alone would affect performance of students, perceived cognitive load, perceived system usability, perceived ease of use, perceived usefulness, system quality, e-learner's satisfaction, willingness to continue the usage and relative advantage in a summative CBA environment.

Table 4.22

Distribution of Gender, age and marks obtained in the last qualifying exam across Exam conditions and the countries.

	Single Question		Question Customization	
	India	**USA**	**India**	**USA**
Male	15	15	13	10
Female	9	17	8	19
Total	24	32	20	29
Age	22.20 (*M*) 2.08 (*SD*)	18.25(*M*) 0.62(*SD*)	23.00(*M*) 3.19(*SD*)	19.00(*M*) 1.98(*SD*)
Marks obtained in last qualifying exam	63.52(*M*) 7.15(*SD*)	70.81(*M*) 9.20(*SD*)	65.82(*M*) 6.10(*SD*)	69.25(*M*) 7.25(*SD*)

Table 4.23

Descriptive statistics of the dependent measures.

	Single Question				Question Customization			
	India		USA		India		USA	
	M	***S.D***	***M***	***S.D***	***M***	***S.D***	***M***	***S.D***
NASATLX	38.54	13.09	42.07	11.19	43.61	10.89	35.17	18.85
SUS	72.69	13.86	79.14	12.77	66.65	18.27	78.88	16.71
PEOU	16.54	2.50	15.25	3.83	13.00	3.32	16.31	4.43
PEU	11.50	2.64	11.16	3.55	9.20	3.25	11.79	3.69
SQ	41.46	7.99	43.50	5.99	40.55	5.26	43.28	8.46
RA	7.12	1.33	5.66	2.06	6.15	1.81	6.38	2.24
PES	48.00	8.48	41.97	13.11	35.55	13.62	45.34	14.39
WCU	7.75	1.87	6.75	2.18	5.95	1.36	7.45	2.06
TM	5.63	1.58	6.31	1.42	5.15	2.28	5.97	1.55
TK	11.12	5.07	8.40	1.83	9.12	3.95	6.58	3.08

Table 4.24

Summarized results of the tests of Normality (Q-Q plots).

Dependent Measures	Single question w/o customization		Question customization	
	India	**USA**	**India**	**USA**
NASA TLX	Slight deviations	Almost Normal	Slight deviations	Slight deviations
SUS	Almost Normal	Almost Normal	Slight deviations	Almost Normal
PEOU	Almost Normal	Slight deviations	Slight deviations	Slight deviations
PEU	Almost Normal	Slight deviations	Slight deviations	Slight deviations
SQ	Almost Normal	Almost Normal	Slight deviations	Almost Normal
RA	Perfectly Normal	Perfectly Normal	Perfectly Normal	Perfectly Normal
WCU	Perfectly Normal	Perfectly Normal	Almost Normal	Almost Normal
PES	Almost Normal	Almost Normal	Slight deviations	Slight deviations
TM	Perfectly Normal	Perfectly Normal	Perfectly Normal	Perfectly Normal
TK	Almost Normal	Almost Normal	Slight deviations	Slight deviations

Table 4.22 and Table 4.23 highlight the descriptive statistics of the participants and the dependent measures that were captured during the experiment. In the single question without customization condition 24 students (15 male and 9 female; M_{Age} = 22.20, SD_{Age} = 2.08; M_{Marks} = 63.52, SD_{Marks} = 7.15) from India participated, while 32 students (15 male and 17 female; M_{Age} = 18.25, SD_{Age} = 0.62; M_{Marks} = 70.81, SD_{Marks} = 9.20) from USA participated. For the question customization condition again 20 students (13 male and 8 female; M_{Age} = 23.00, SD_{Age} = 3.19; M_{Marks} = 65.82, SD_{Marks} = 6.10) from India participated, while 29 students (10 male and 19 female; M_{Age} = 19.00, SD_{Age} = 1.98; M_{Marks} = 69.25, SD_{Marks} = 7.25) from USA participated. All the four groups were homogeneous in terms of their marks obtained in the last qualifying exam but not in terms of their age (See Appendix 4, page 240).

The inferential statistical analysis was carried out starting first with the tests of Normality. Q-Q plots were generated to test the normality of the data distribution across the groups for all the dependent measures. **Appendix-3D** enlists the Q-Q plots of the data distributions. The observations are summarized in table 4.24. The Q-Q plots revealed that there are slight deviations from Normality for some of the dependent measures. Thereafter, the assumption of homogeneity of variance-covariances were tested using Box's Test. Box's test of equality of covariance matrices highlights that it is significant at p<0.005. The assumption of homogeneity of variance-covariance matrices was violated. Therefore, the Pillai's statistic was used to derive the conclusions from the multivariate tests as it is robust to the violations of these assumptions (Haase & Ellis, 1987).

Table 4.25

Multivariate statistics

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Observed Power[c]
Exam	**Pillai's Trace**	0.15	1.68	10.00	92.00	0.10	0.15	0.77
Country	**Pillai's Trace**	0.30	3.88	10.00	92.00	0.00	0.30	0.99
Exam * Country	**Pillai's Trace**	0.20	2.25	10.00	92.00	0.02	0.20	0.90

Table 4.26

Univariate statistics of the dependent measures based on the independent factor – country and the interaction of the factors.

Source	Dependent Variable	Type III Sum of Squares	*df*	Mean Square	*F*	Sig.	Partial Eta Squared	Observed Power[k]
Country	**NASATLX**	152.65	1	152.65	0.77	0.38	0.01	0.14
	SUS	2217.50	1	2217.50	9.48	0.00	0.09	0.86
	PEOU	25.89	1	25.89	1.93	0.17	0.02	0.28
	PEU	32.14	1	32.14	2.87	0.09	0.03	0.39
	SQ	144.41	1	144.41	2.85	0.09	0.03	0.39
	RA	9.76	1	9.76	2.63	0.11	0.03	0.36
	PES	89.99	1	89.99	0.56	0.46	0.01	0.11
	WCU	1.58	1	1.58	0.42	0.52	0.00	0.10
	TM	14.35	1	14.35	5.07	0.03	0.05	0.61
	TK	175.46	1	175.46	14.08	0.00	0.12	0.96
Exam * Country	**NASATLX**	910.06	1	910.06	4.59	0.03	0.04	0.56
	SUS	211.97	1	211.97	0.91	0.34	0.01	0.16
	PEOU	134.55	1	134.55	10.01	0.00	0.09	0.88
	PEU	54.80	1	54.80	4.89	0.03	0.05	0.59
	SQ	2.97	1	2.97	0.06	0.81	0.00	0.06
	RA	18.32	1	18.32	4.93	0.03	0.05	0.59
	PES	1591.27	1	1591.27	9.86	0.00	0.09	0.87
	WCU	39.65	1	39.65	10.47	0.00	0.09	0.89
	TM	0.10	1	0.10	0.04	0.85	0.00	0.05
	TK	0.22	1	0.22	0.02	0.89	0.00	0.05

The 2x2 factorial multivariate tests (MANOVA) conducted on the data sets (Table 4.25) revealed that there was a statistically significant main effect for Country , F (10, 92) = 3.88, $p < 0.001$; Pillai's Trace = 0.30; partial $\eta^2 = 0.30$. There was also a statistically significant interaction effect of Country and interface, F (10, 92) = 2.25, $p < 0.05$; Pillai's Trace = 0.20; partial $\eta^2 = .20$. There was no statistically significant main effect for interface feature.

It is important to state here that because of the observed significant interaction effect between cultural background (country) and interface, the main effects of these criterion variables which have been found significant no longer demands importance. Because it is evident from the significant interaction effect that one level of the criterion variable depends on the level of the other criterion variable to affect the group measures. Therefore, after the significant interaction MANOVA effect, separate univariate tests was carried out to figure out the dependent measures which account for significant differences across the groups.

Univariate analysis conducted reveals that the group membership (Table 4.26), the interacting criterion variables interface feature and cultural background had a significant effect on subjective cognitive load, $F(1, 101) = 4.59$, $p < 0.05$; partial $\eta^2 = 0.04$; perceived ease of use, $F(1, 101) = 10.01$, $p < 0.001$; partial $\eta^2 = 0.09$; perceived usefulness, $F(1, 101) = 4.89$, $p < 0.05$; partial $\eta^2 = 0.05$; relative advantage, $F(1, 101) = 4.93$, $p < 0.05$; partial $\eta^2 = 0.05$; perceived e-learner satisfaction $F(1, 101) = 9.86$, $p < 0.001$; partial $\eta^2 = 0.09$ and willingness to continue to use, $F(1, 101) = 10.47$, $p < 0.001$; partial $\eta^2 = 0.09$.Though, it was clear from the separate univariate tests that group differences among subjective cognitive load, perceived ease of use, perceived usefulness , relative advantage, perceived e-learner's satisfaction and willingness to continue to use can be explained by the interacting criterion variables but it further analysis is required regarding which particular groups differ significantly.

To identify the significant differences among groups separate independent sample *t*-tests were employed. Appendix- 3D enlists the independent sample *t*-test tables. It is important to state here that *t*-tests are generally robust against violations of normality (Edgell & Noon, 1984). After *t*-tests were conducted on with single question without customization among the Indian and the American students it revealed that there was a significant difference in the perceived system usability scores in the single question without customization condition for Indian students ($M = 72.68$, $SD = 13.85$) and American students ($M = 79.14$, $SD = 12.77$); t (54) = -1.80, $p = 0.035$ (single tailed), d= -0.49. A significant difference in the relative advantage scores in the single question without customization condition for Indian students ($M = 7.12$, $SD = 1.32$) and American students ($M = 5.65$, $SD = 2.05$); t (54) = 3.04, $p = 0.004$ (two tailed), d=

0.82. A significant difference in the perceived e-learner's satisfaction scores in the single question without customization condition for Indian students (M = 48.00, SD = 8.48) and American students (M = 41.96, SD = 13.11); t (54) = 1.964, p = 0.027 (single tailed), d= 0.53. A significant difference in the willingness to continue to use scores in the single question without customization condition for Indian students (M =7.75, SD =1.87) and American students (M =6.75, SD =2.18); t (54) =1.80, p =0.037 (one tailed), d= 0.48. A significant difference in the total marks in the single question without customization condition for Indian students (M =5.62, SD =1.58) and American students (M =6.31, SD =1.42); t (54) = -1.70, p =0.047 (one tailed), d= -0.46. A significant difference in the total time taken to complete the test in the single question without customization condition for Indian students (M =11.12, SD =5.07) and American students (M =8.40, SD =1.83); t (54) = 2.80, p =0.007 (two tailed), d= 0.76.

The *t*-tests on the question customization CBA condition dependent measures among the Indian and the American students revealed that there is a significant difference in the subjective cognitive load scores for Indian students (M =43.60, SD =10.88) and the American students (M =35.17, SD =18.85); t(47)= 1.80, p = 0.039 (single tailed), d = 0.52. A significant difference in the perceived system usability scores in the question customization condition for Indian students (M =66.65, SD =18.26) and the American students (M =78.87, SD =16.71); t (47) = -2.420, p = 0.019 (two tailed), d = -0.70. A significant difference in the perceived ease of use scores in the question customization condition for Indian students (M =13.00, SD =3.32) and the American students (M =16.31, SD = 4.43); t (47) = -2.832, p = 0.007 (two tailed), d = -0.82. A significant difference in the perceived usefulness scores in the *question customization* condition for Indian students (M =9.20, SD =3.25) and the American students (M =11.79, SD =3.68); t (47) = -2.535, p = 0.015 (two tailed), d = -0.74. A significant difference in the perceived e-learner's satisfaction scores in the *question customization* condition for Indian students (M =35.55, SD =13.61) and the American students (M = 45.34, SD =14.38); t (47) = -2.393, p = 0.021 (two tailed), d = -0.69.A significant difference in the willingness to continue to use scores in the *question customization* condition for Indian students (M =5.95, SD =1.35) and the American students (M =7.44, SD =2.06); t (47) = -2.847, p = 0.007 (two tailed), d = -0.83. A significant difference in the total time taken to complete the assessment session in the *question customization* condition for Indian students (M =9.11, SD = 3.94) and the American students (M =6.58, SD = 3.07); t (47) = 2.522, p = 0.015 (two tailed), d = 0.73.

Thereafter, *t* tests were conducted on the dependent measures among the same student communities for both the CBA conditions. A significant difference in the perceived ease of use scores for the Indian students were observed for the *single question without customization* condition (M =16.54, SD = 2.50) and *question customization* CBA condition (M =13.00, SD = 3.32); t (42) = 4.02, p = 0.000 (two tailed), d = 1.24. A significant difference in the perceived

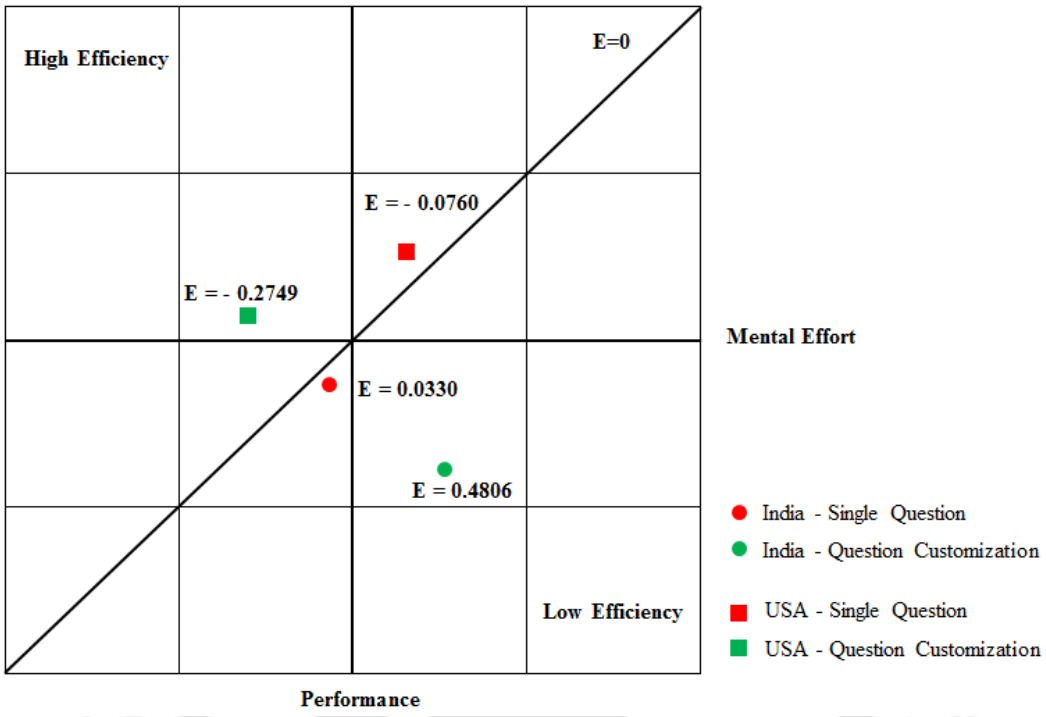


Figure 4.24. Instructional efficiency of Single question and Question customization CBA conditions between Indian and American students.

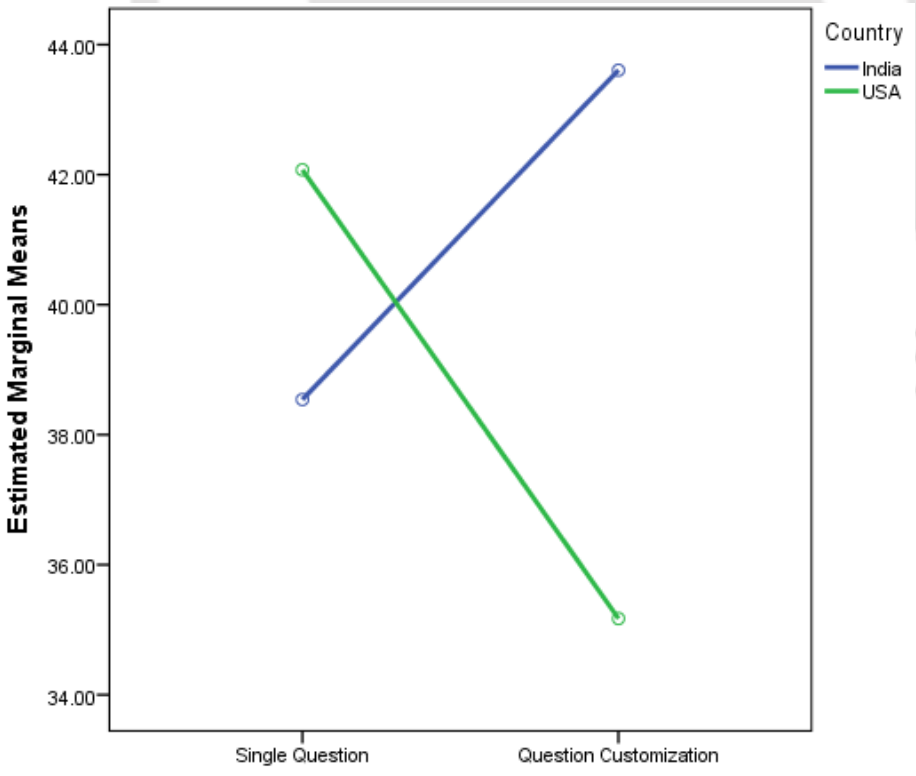


Figure 4.25. Estimated Marginal means of NASA TLX across the test conditions and countries.

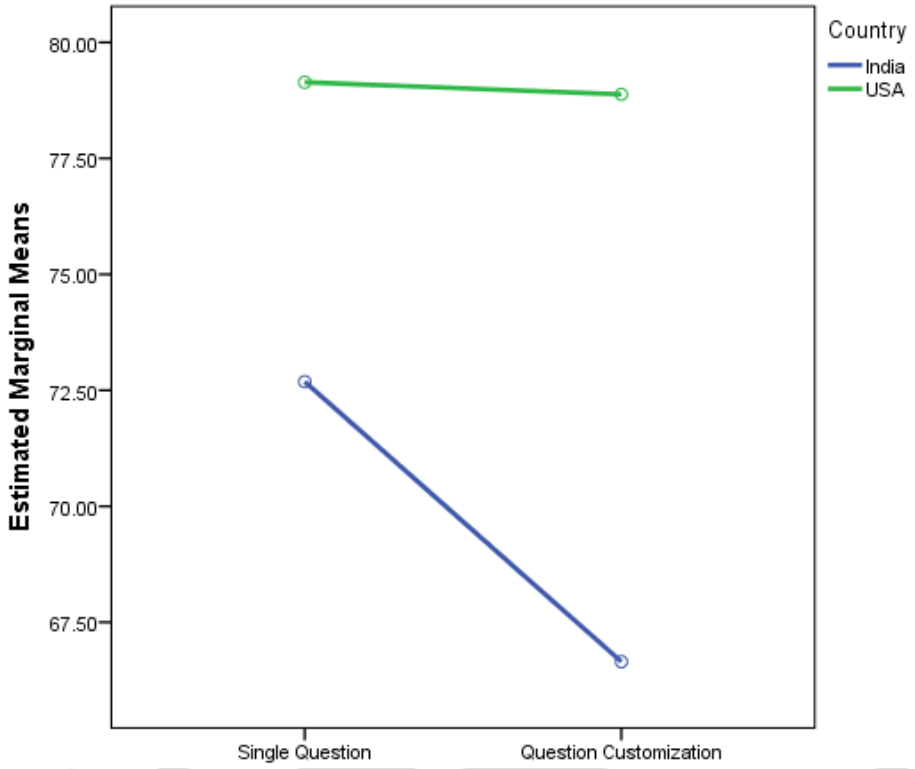


Figure 4.26. Estimated Marginal means of SUS across the test conditions and countries.

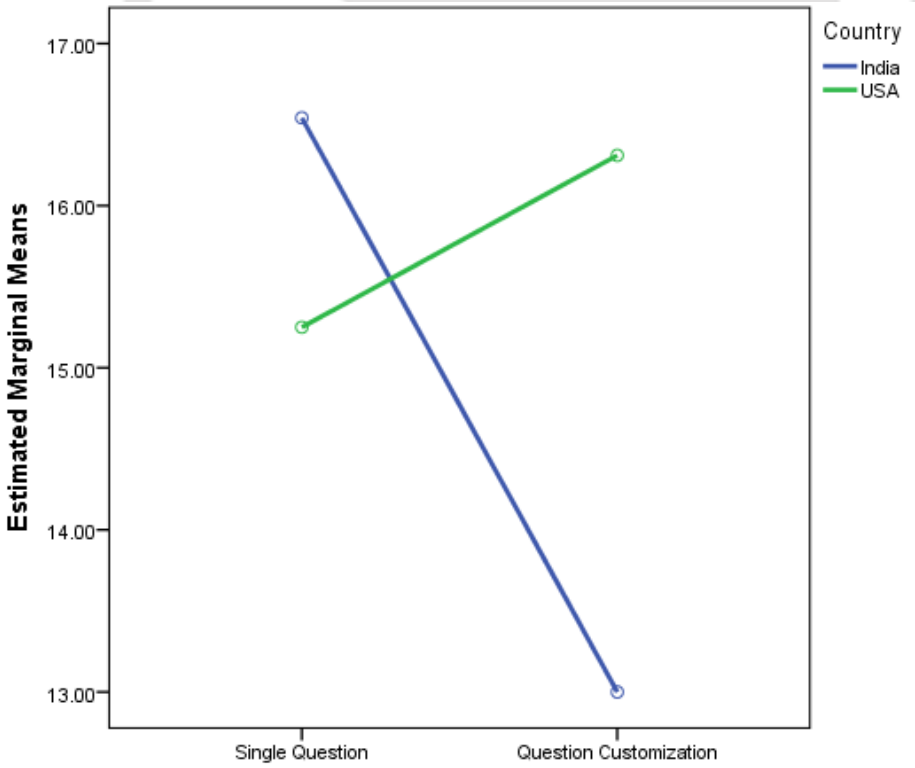


Figure 4.27. Estimated Marginal means of PEOU across the test conditions and countries.

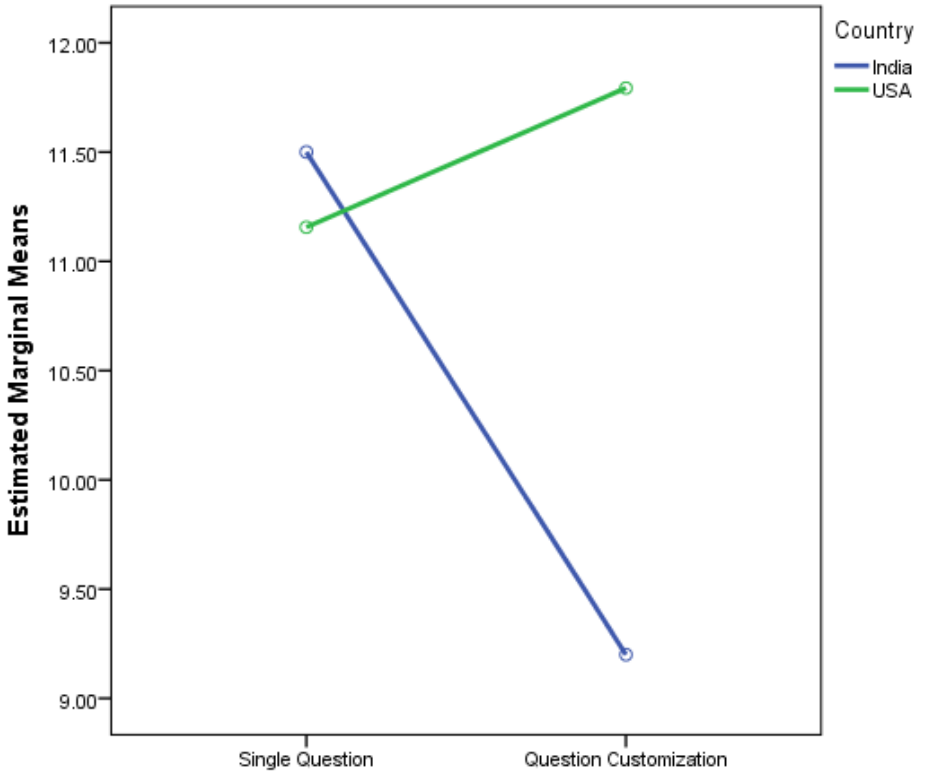


Figure 4.28. Estimated Marginal means of PEU across the test conditions and countries.

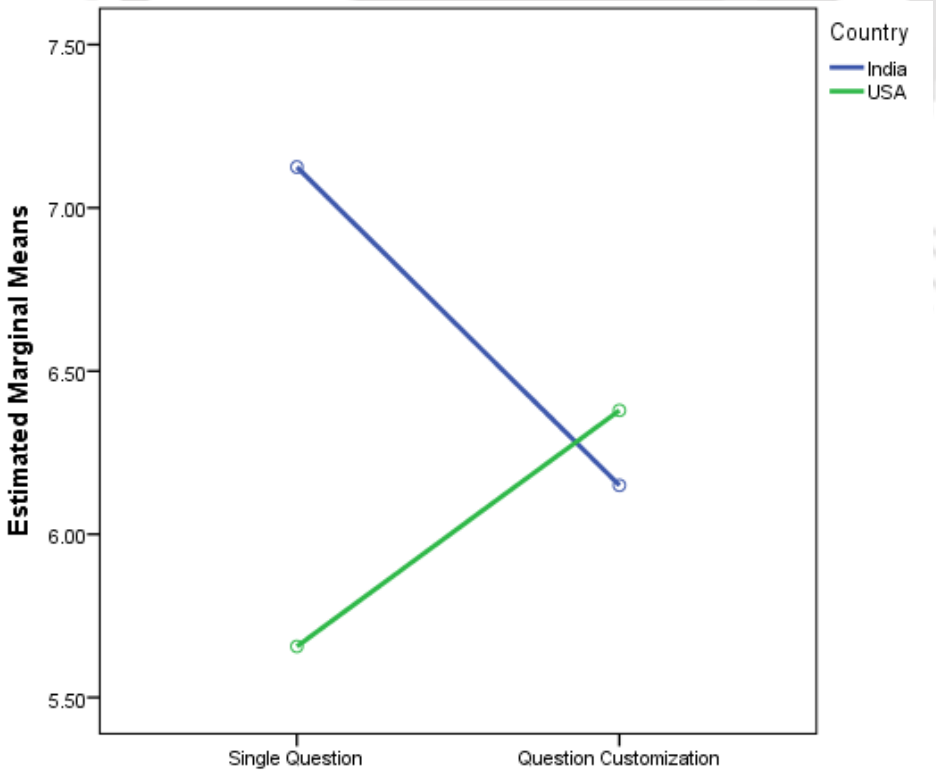


Figure 4.29. Estimated Marginal means of RA across the test conditions and countries.

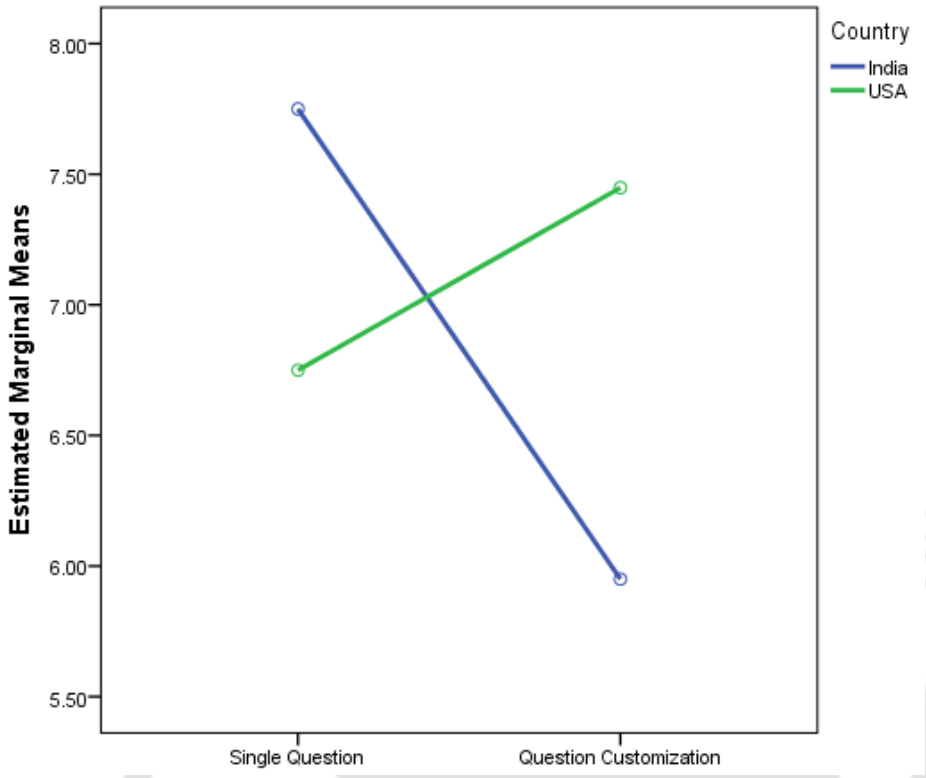


Figure 4.30. Estimated Marginal means of WCU across the test conditions and countries.

usefulness scores for the Indian students were observed for the *single question without customization* condition (*M* =11.50, *SD* = 2.63) and *question customization* CBA condition (*M* = 9.20, *SD* = 3.25); *t* (42) = 2.59, *p* = 0.013 (two tailed), *d* = 0.79. A significant difference in the relative advantage scores for the Indian students were observed for the *single question without customization* condition (*M* =7.12, *SD* = 1.32) and *question customization* CBA condition (*M* = 6.15, *SD* = 1.81); *t* (42) = 2.055, *p* = 0.046 (two tailed), *d* = 0.63.

A significant difference in the perceived e-learner's satisfaction scores for the Indian students were observed for the *single question without customization* condition (*M* =48.00, *SD* = 8.48) and *question customization* CBA condition (*M* = 35.55, *SD* = 13.61); *t* (42) = 3.704, *p* = 0.001 (two tailed), *d* = 1.14. A significant difference in the willingness to continue to use scores for the Indian students were observed for the *single question without customization* condition (*M* = 7.75, *SD* = 1.87) and *question customization* CBA condition (*M* = 5.95, *SD* = 1.35); *t* (42) = 3.586, *p* = 0.001 (two tailed), *d* = 1.10.

Table 4.27

Hypotheses testing result

Hypothesis $\mathbf{H_{4b}}$ = Culture affects	Supported
Subjective Cognitive Load	Not Supported
Perceived System Usability	Supported
Perceived Ease of Use	Not Supported
Perceived Usefulness	Not Supported
Perceived System Quality	Not Supported
Relative Advantage	Not Supported
Perceived E-learner's Satisfaction	Not Supported
Willingness to Continue to Use	Not Supported
Total Marks	Supported
Time Taken	Supported
Hypothesis $\mathbf{H_{4c}}$ = Cultural background X Interface feature affects	Supported
Subjective Cognitive Load	Supported
Perceived System Usability	Not Supported
Perceived Ease of Use	Supported
Perceived Usefulness	Supported
Perceived System Quality	Not Supported
Relative Advantage	Supported
Perceived E-learner's Satisfaction	Supported
Willingness to Continue to Use	Supported
Total Marks	Not Supported
Time Taken	Not Supported
Hypothesis $\mathbf{H_{4a}}$ = Interface feature affects	Not Supported

For the American students, a significant difference in the subjective cognitive load scores were observed for the *single question without customization* condition (M = 42.07, SD = 11.19) and *question customization* CBA condition (M = 35.17, SD = 18.85); t (59) = 1.758, p = 0.042 (single tailed), d = 0.45. A significant difference in the total time taken was observed for American students in the *single question without customization* condition (M = 8.40, SD = 1.83)

and *question customization* CBA condition (M = 6.58, SD = 3.07); t (59) = 2.833, p = 0.006 (two tailed), d = 0.73.

The instructional efficiency plotted in the Cartesian coordinate system (Figure 4.24) highlights that for the Indian students both the CBA conditions fall in the low instructional efficiency zone, while for the American students both the CBA conditions fall in the high instructional efficiency zone. It is important to highlight here that, though for both the student communities the instructional efficiency of the two CBA conditions fall in the same zone, one of them has slight edge over the other in terms of the instructional efficiency view. While, for the American students *question customization* seems to have higher efficiency than *single question without customization* based CBA system; but for the Indian students *single question without customization* is relatively more efficient than the *question customization* CBA condition. Figure 4.25-4.30 highlights the marginal means of the dependent measures for the CBA conditions across the two student communities.

Discussion: Experiment 4

The results of the experiment highlight that there is a significant interaction of interface features and cultural background on the dependent measures subjective cognitive load, perceived ease of use, perceived usefulness, relative advantage, perceived e-learner satisfaction and willingness to continue to use. The significant interaction effect highlights that one level of the interface feature depends on the level of cultural backgrounds to affect these dependent measures. The result of the hypotheses testing is listed in table 4.27 and the summary of the results have been reported in the table 4.28(a) and 4.28(b). The results points out to some interesting facts. It has been observed that overall Indian students prefer *single question without customization* feature based CBA systems while American students prefer *question customization* feature based CBA systems. American students report low subjective cognitive load in *question customization* in comparison to the *single question without customization* feature based CBA systems. Overall American students report low subjective cognitive load in comparison to their Indian counterparts for the *question customization* feature based CBA system. For the Indian students perceived ease of use of *single question without customization* was higher in comparison to *question customization* feature based CBA system. Similarly for Indian students perceived usefulness of *question customization* was lower in comparison to *single question without customization* feature based CBA system. For Indian students' perceived e-learners' satisfaction in *single question without customization* was higher in comparison to *question customization* feature based CBA system. For Indian students relative advantage of *single question without customization* was higher in comparison to *question*

Table 4.28 (a)

Summary of Results

Dependent Measures	**Results**
Subjective Cognitive Load	Overall the Indian students rated 'cognitive load' of question customization significantly higher than their American counterparts. American students reported 'subjective cognitive load' of single question without customization significantly higher than the question customization CBA condition.
Perceived System Usability	Overall the American students rated 'system usability' of single question without customization significantly higher than their Indian counterparts. Overall the Indian students rated system usability of question customization significantly lower than their American counterparts.
Perceived Ease of Use	Overall the Indian students rated 'ease of use' of question customization significantly lower than their American counterparts. Indian students rated 'ease of use' of single question without customization significantly higher than the question customization CBA condition.
Perceived Usefulness	Overall the Indian students rated 'perceived usefulness' of question customization significantly lower than their American counterparts. Indian students rated 'perceived usefulness' of single question without customization significantly higher than the question customization CBA condition.
Perceived e-learner's satisfaction	Overall the American students rated 'perceived e-learner satisfaction' of single question without customization significantly lower than their Indian counterparts. Overall the Indian students rated 'perceived e-learner's satisfaction' of question customization significantly lower than their American counterparts. Indian students rated 'perceived e-learner's satisfaction' of single question without customization significantly higher than the question customization CBA condition.

Table 4.28 (b)

Summary of Results

Dependent Measures	**Results**
Relative Advantage	Overall the American students rated 'relative advantage' of single question without customization significantly lower than their Indian counterparts. Indian students rated 'relative advantage' of single question without customization significantly higher than the question customization CBA condition.
Willingness to continue to use	Overall the American students rated 'willingness to continue to use' of single question without customization significantly lower than their Indian counterparts. Overall the Indian students rated 'willingness to continue to use' of question customization significantly lower than their American counterparts. Indian students rated 'willingness to continue to use' of single question without customization significantly higher than the question customization CBA condition.
Total Marks	Overall the American students scored significantly higher marks in single question without customization than their Indian counterparts.
Time Taken	Overall the American students took significantly lower time in single question without customization than their Indian counterparts. Overall the American students took significantly lower time in question customization than their Indian counterparts. American students took significantly more time in single question without customization than the question customization CBA condition.

customization feature based CBA system. For American students it was just the opposite. Indian students' willingness to continue to use the *system for single question without customization* was higher in comparison to *question customization* feature based CBA system. American students took significantly longer time to complete the assessment session in *single question without customization* in comparison to *question customization* feature based CBA system.

The instructional efficiency plot (Figure 4.24) highlights that for the Indian students both the CBA conditions fall in the low instructional efficiency zone, while for the American students both the CBA conditions fall in the high instructional efficiency zone. For the American students question customization seems to have higher efficiency than *single question without customization*; but for the Indian students' *single question without customization* is relatively more efficient than the *question customization* CBA condition.

4.2.5 Experiment 5: Low screen density (maximum white space) vs. High screen density (minimum white space)

The working hypotheses which were tested in this experiment Low screen density (maximum white space) versus High screen density (minimum white space) are as follows:

H_{5a}: White space or screen density alone would affect performance of students, perceived cognitive load, perceived system usability, perceived ease of use, perceived usefulness, system quality, e-learner's satisfaction, willingness to continue the usage and relative advantage in a summative CBA environment.

H_{5b}: Cultural background alone would affect performance of students, perceived cognitive load, perceived system usability, perceived ease of use, perceived usefulness, system quality, e-learner's satisfaction, willingness to continue the usage and relative advantage in a summative CBA environment.

H_{5c}: The interacting effect of cultural background and white space or screen density alone would affect performance of students, perceived cognitive load, perceived system usability, perceived ease of use, perceived usefulness, system quality, e-learner's satisfaction, willingness to continue the usage and relative advantage in a summative CBA environment.

Table 4.29 and table 4.30 highlight the descriptive statistics of the participants and the dependent measures that were captured during the experiment. In the low screen density (maximum white space) condition 28 students (17 male and 11 female; M_{Age} = 23.24, SD_{Age} = 2.15; M_{Marks} = 72.32, SD_{Marks} = 6.00) from India participated, similarly 28 students (10 male and 18 female; M_{Age} = 18.50, SD_{Age} = 1.10; M_{Marks} = 70.10, SD_{Marks} = 7.50) from USA participated. For the high screen density (minimum white space) condition again 30 students (20 male and 10 female; M_{Age} = 22.15, SD_{Age} = 2.00; M_{Marks} = 73.15, SD_{Marks} = 6.55) from India participated, while 27 students (10 male and 17 female; M_{Age} = 19.54, SD_{Age} = 0.91; M_{Marks} = 69.55, SD_{Marks} = 8.45) from USA participated. All the four groups were homogeneous in terms of their marks obtained in the last qualifying exam but not in terms of their age (See Appendix 4, page 240).

Table 4.29

Distribution of Gender, age and marks obtained in the last qualifying exam across Exam conditions and the countries.

	Low Screen Density		High Screen Density	
	India	**USA**	**India**	**USA**
Male	17	10	20	10
Female	11	18	10	17
Total	28	28	30	27
Age	23.24 (*M*) 2.15 (*SD*)	18.50(*M*) 1.10(*SD*)	22.15(*M*) 2.00(*SD*)	19.54(*M*) 0.91(*SD*)
Marks obtained in last qualifying exam	72.32(*M*) 6.00(*SD*)	70.10(*M*) 7.50(*SD*)	73.15(*M*) 6.55(*SD*)	69.55(*M*) 8.45(*SD*)

Table 4.30

Descriptive statistics of the dependent measures.

	Low Screen Density				High Screen Density			
	USA		India		USA		India	
	M	***S.D***	***M***	***S.D***	***M***	***S.D***	***M***	***S.D***
NASA TLX	30.21	14.51	32.88	16.53	33.75	13.48	36.43	17.61
SUS	85.09	11.76	70.45	15.02	81.02	10.81	68.92	13.48
PEOU	17.21	2.96	17.25	2.46	17.93	2.48	16.30	2.95
PEU	12.46	2.67	11.82	3.02	12.81	1.92	11.50	2.61
SQ	44.18	6.28	42.75	8.25	46.70	4.56	43.10	12.48
RA	6.61	1.64	7.18	2.06	6.19	1.62	7.23	1.74
PES	48.11	11.93	46.68	10.79	48.56	8.88	46.13	10.08
WCU	7.32	2.04	7.11	2.33	7.44	1.99	7.20	1.61
TM	6.36	.95	5.82	1.28	6.52	1.28	5.53	1.61
TK	7.13	2.03	7.27	4.16	5.02	1.58	11.50	23.54

Table 4.31

Summarized results of the tests of Normality (Q-Q plots).

Dependent Measures	Low screen density		High screen density	
	India	**USA**	**India**	**USA**
NASA TLX	Almost Normal	Almost Normal	Almost Normal	Almost Normal
SUS	Perfectly Normal	Almost Normal	Perfectly Normal	Almost Normal
PEOU	Almost Normal	Slight deviations	Almost Normal	Slight deviations
PEU	Slight deviations	Slight deviations	Slight deviations	Slight deviations
SQ	Almost Normal	Almost Normal	Almost Normal	Almost Normal
RA	Almost Normal	Perfectly Normal	Perfectly Normal	Almost Normal
WCU	Slight deviations	Slight deviations	Slight deviations	Almost Normal
PES	Almost Normal	Almost Normal	Almost Normal	Almost Normal
TM	Almost Normal	Almost Normal	Almost Normal	Almost Normal
TK	Slight deviations	Slight deviations	Slight deviations	Slight deviations

The inferential statistical analysis was carried out starting first with the tests of Normality. Q-Q plots were generated to test the normality of the data distribution across the groups for all the dependent measures. **Appendix-3E** enlists the Q-Q plots of the data distributions. The observations are summarized in table 4.31. The Q-Q plots revealed slight deviations from Normality for some of the dependent measures. Thereafter, the assumption of homogeneity of variance-covariances were tested using Box's Test. Box's test of equality of covariance matrices highlights that it is significant at $p<0.005$. The assumption of homogeneity of variance-covariance matrices was violated. Therefore to draw inferences from the multivariate tests the Pillai's statistic is used as it is robust to the violations of these assumptions (Haase & Ellis, 1987).

The 2x2 factorial multivariate tests (MANOVA) conducted on the data sets (Table 4.32) revealed that there was a statistically significant main effect for cultural background, F (10, 96) = 4.00, $p < .001$; Pillai's Trace = 0.29; partial η^2 = 0.29. There was no statistically significant main effect for Interface feature neither a statistically significant interaction effect between Country and interface feature.

Table 4.32

Multivariate statistics

Effect		Value	*F*	Hypothesis df	Error df	Sig.	Partial Eta Squared	Observed Power[c]
Exam	**Pillai's Trace**	0.06	0.64[b]	10.00	100.00	0.77	0.06	0.32
Country	**Pillai's Trace**	0.36	5.74[b]	10.00	100.00	0.00	0.36	1.00
Exam * Country	**Pillai's Trace**	0.09	1.00[b]	10.00	100.00	0.45	0.09	0.50

Table 4.33

Univariate statistics of the dependent measures based on the independent factor – country.

Source	Dependent Variable	Type III Sum of Squares	*df*	Mean Square	*F*	Sig.	Partial Eta Squared	Observed Power[k]
Country	**NASA TLX**	202.51	1	202.51	0.82	0.37	0.01	0.15
	SUS	5044.32	1	5044.32	30.33	0.00	0.22	1.00
	PEOU	17.83	1	17.83	2.39	0.12	0.02	0.34
	PEU	27.03	1	27.03	4.02	0.05	0.04	0.51
	SQ	178.59	1	178.59	2.44	0.12	0.02	0.34
	RA	18.50	1	18.50	5.89	0.02	0.05	0.67
	PES	104.57	1	104.57	0.95	0.33	0.01	0.16
	WCU	1.48	1	1.48	0.37	0.54	0.00	0.09
	TM	16.31	1	16.31	9.52	0.00	0.08	0.86
	TK	310.01	1	310.01	2.02	0.16	0.02	0.29

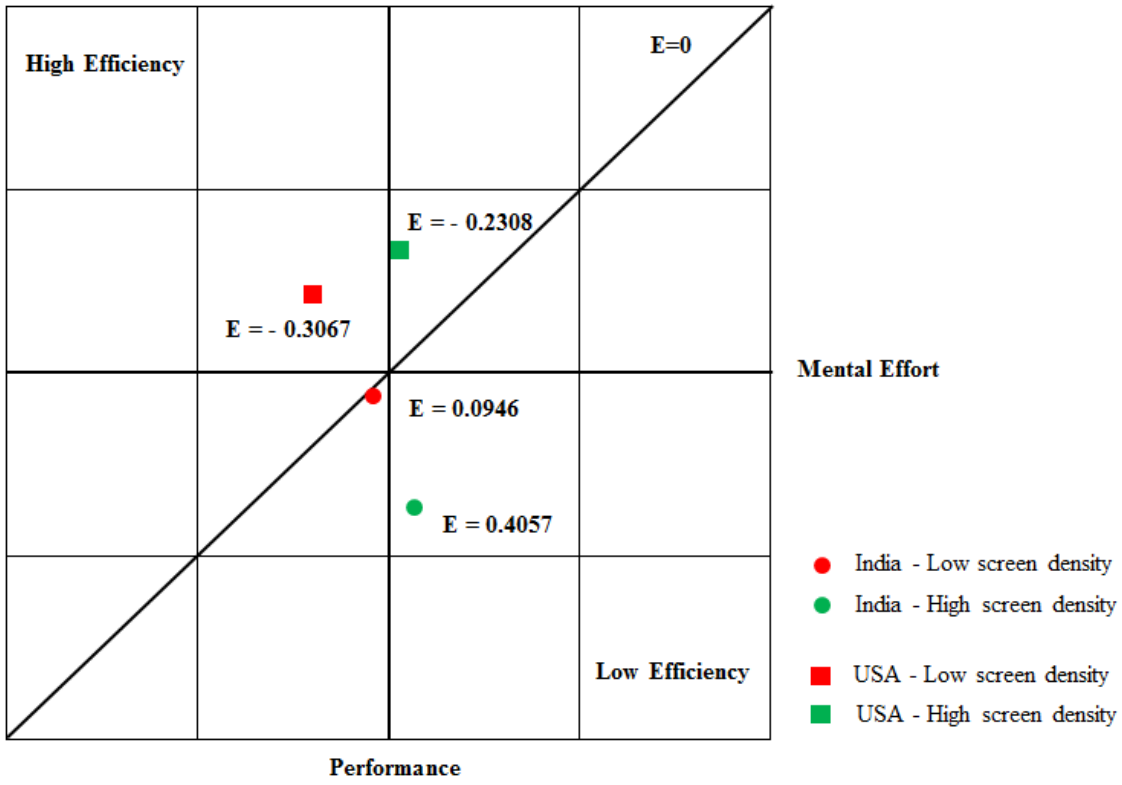


Figure 4.31. Instructional efficiency of Low screen density (maximum white space) and high screen density (minimum white space) CBA conditions between Indian and American students.

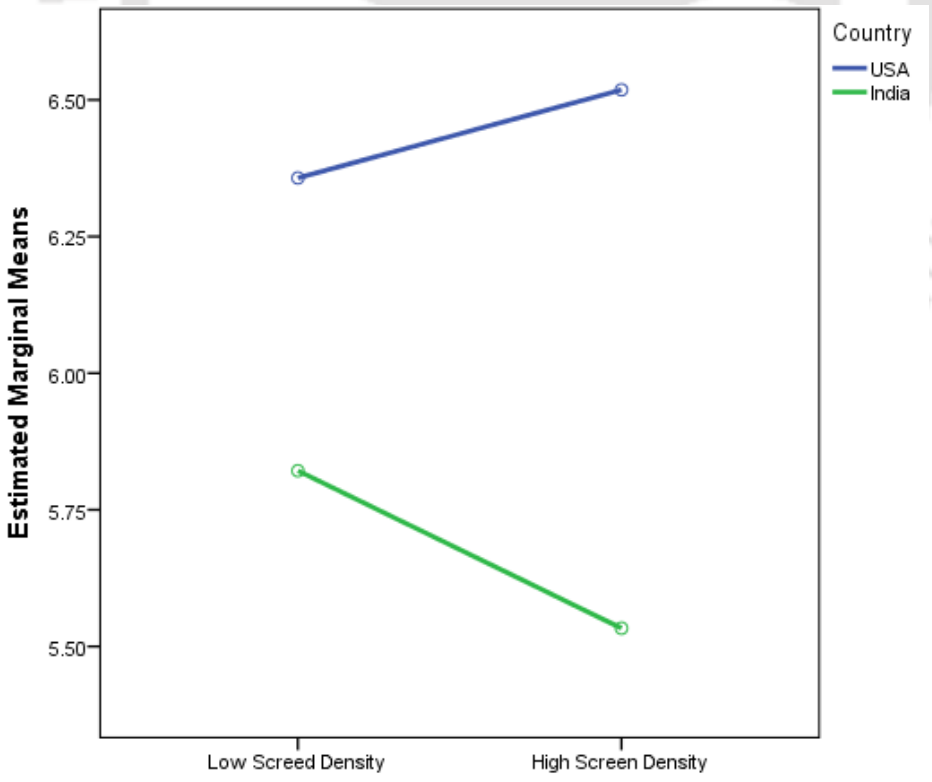


Figure 4.32. Estimated Marginal means of TM across the test conditions and countries.

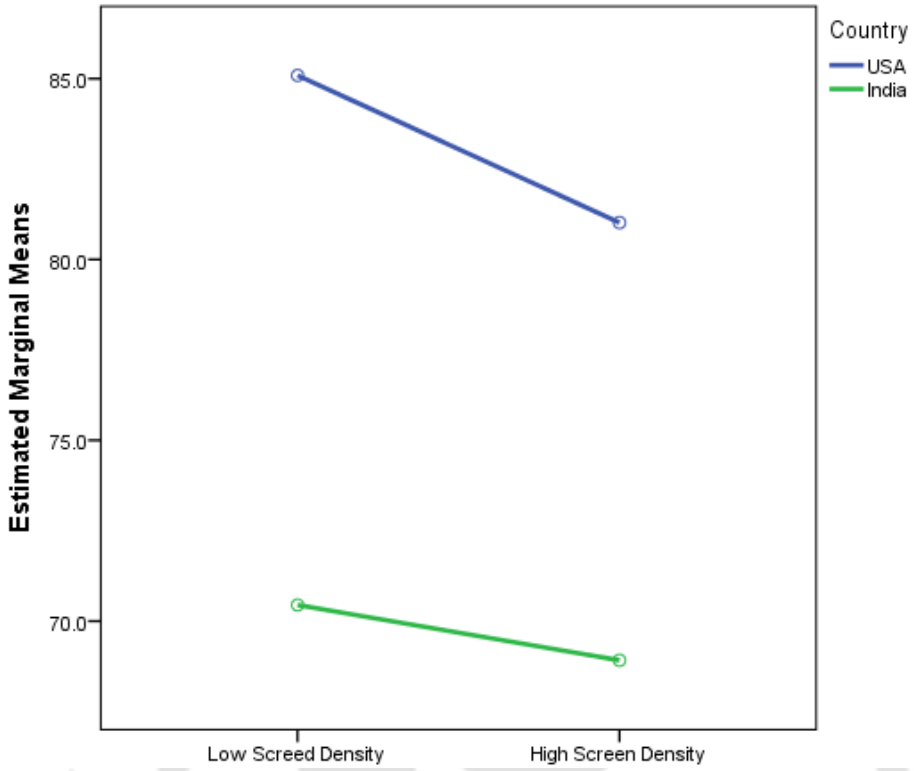


Figure 4.33. Estimated Marginal means of SUS across the test conditions and countries.

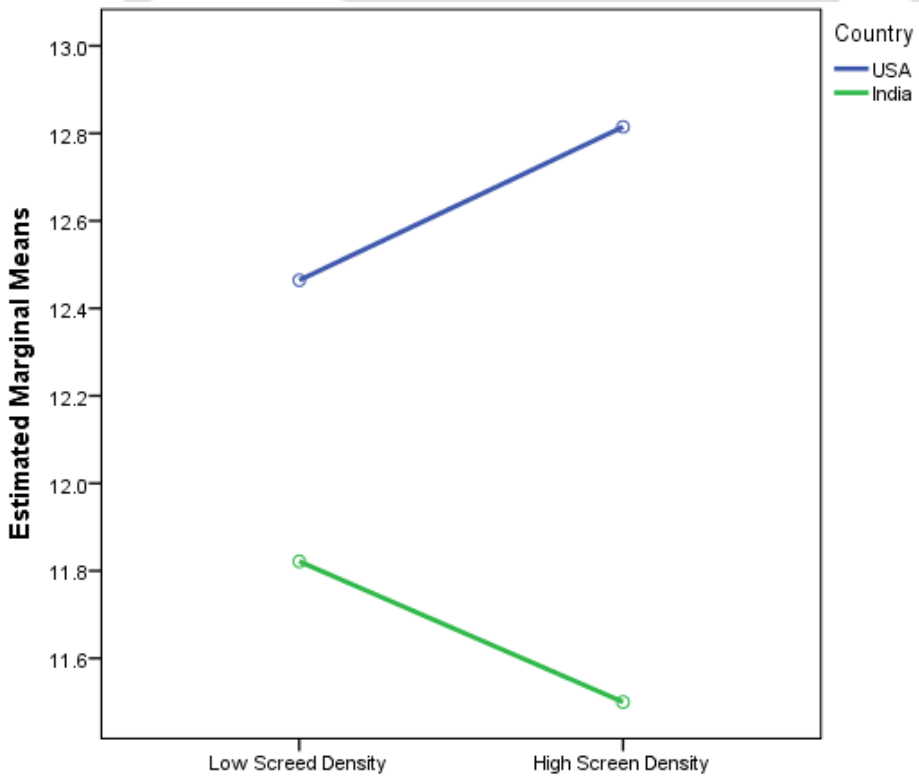


Figure 4.34. Estimated Marginal means of PEU across the test conditions and countries.

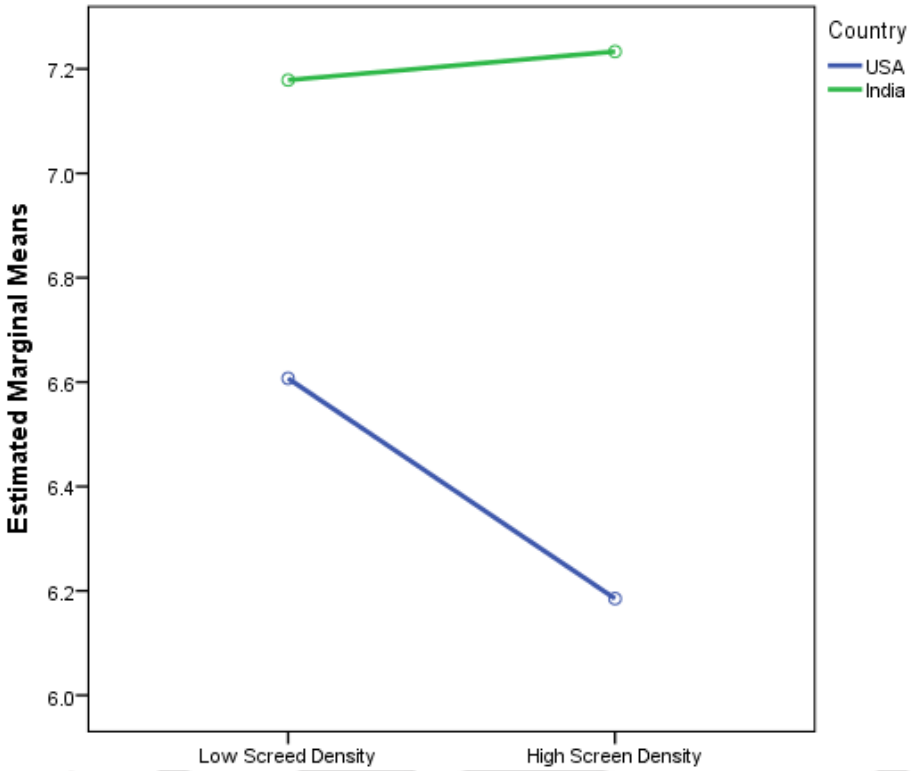


Figure 4.35. Estimated Marginal means of RA across the test conditions and countries.

Univariate analysis conducted reveals that the group membership (Table 4.33), Cultural background had a significant effect on perceived system usability, F (1, 109) = 30.33, p < 0.001; partial η^2 = 0.22; relative advantage, F (1, 109) = 5.89, p < 0.05; partial η^2 = 0.05 and total marks, F (1, 109) = 9.52, p < 0.001; partial η^2 = 0.08.

The pair wise analysis conducted to identify the direction of the effect highlights that American students have reported significantly high system usability scores for both the CBA conditions in comparison to their Indian counterparts. Indians on the other hand reported significantly high relative advantage scores for both the CBA conditions in comparison to their American counterparts. American students score significantly higher in both the CBA conditions in comparison to their Indian counterparts. Figure 4.31 highlights the instructional efficiency plotted in the Cartesian coordinate system for both the student communities across the CBA conditions. It is interesting to observe that while both the CBA conditions are in high instructional efficiency zone for the American students, for the Indian students both the CBA conditions are in low instructional efficiency zone but among these two CBA conditions for Indians the low white space seems to be comparatively of low instructional

Table 4.34

Hypotheses testing

Hypothesis $\mathbf{H_{5b}}$ = Culture affects	Supported
Subjective Cognitive Load	Not Supported
Perceived System Usability	Supported
Perceived Ease of Use	Not Supported
Perceived Usefulness	Not Supported
Perceived System Quality	Not Supported
Relative Advantage	Supported
Perceived E-learner's Satisfaction	Not Supported
Willingness to Continue to Use	Not Supported
Total Marks	Supported
Time Taken	Not Supported
Hypothesis $\mathbf{H_{5c}}$ = Cultural background X Interface feature affects	Not Supported
Hypothesis $\mathbf{H_{5a}}$ = Interface feature affects	Not Supported

Table 4.35

Summary of results

Dependent Measures	**Results**
Perceived System Usability	Overall American students rated higher 'perceived system usability' than their Indian counterparts.
Relative Advantage	Overall Indian students rated higher 'relative advantage' than their American counterparts.
Total Marks	American students scored comparatively higher marks than the Indian students.

efficiency than the *high white space* CBA condition. Figure 4.32 -4.35 highlights the marginal means of the CBA conditions across the two students communities.

Discussion: Experiment 5

The results of the experiment on the effect of *low screen density* (high white space) vs. *high screen density* (low white space) on performance and the subjective evaluation of the CBA system by the students from India and USA has found that students across the two cultural backgrounds does not get affected due to the two CBA conditions. Table 4.34 highlights the hypotheses testing results and table 4.35 enlists the summary of the findings. The results highlight that students are not concerned to a greater extent with the reduction in the *white space* during CBA sessions. It is important to highlight that Indian students though had a slight preference for the *low screen density* CBA condition but this conclusion may be contended because of the low effect size of the treatment conditions.

4.3. Chapter 4: Summary

This chapter presented the experimental results of the research investigation being argued in this thesis. A 2x2 factorial multivariate analysis has been carried out and the treatment effects on the student communities across India and America has been highlighted. The next chapter discusses the consolidated findings of this research investigation, highlights the contribution of the research work and presents future scope of the current investigation.

Chapter 5

Conclusion

Chapter abstract: *This chapter presents the consolidated findings of the research experiments. It highlights the contribution of this research work and enlists its future scope. Results of five experiments have been presented in chapter 4. A detailed discussion on the results of the experiments followed thereafter. Here, the results of the experiments are presented in the light of the research questions that were raised earlier. The practical and theoretical implications of the results are presented and its limitations highlighted. Finally, the future scope of the research investigation has been discussed in detail.*

5.1. Introduction

This research investigation was carried out on noticing the contradictory findings and the various questions that remained unanswered in 'test mode effect' studies. The literature survey points out to the existence of inconclusive research from the perspective of identifying the effect of GUIs on the individual characteristics of the Examinees. Though it has been argued in the literature that the individual characteristics can be a predictor of test mode effect between critical individual characteristics' that might affect students' performance in a computer based assessment environment had not been found to be addressed earlier.

The "**presentation format effect**" is the focal point around which the research investigation has been carried out in this thesis. It is argued here that test scores which has been cited extensively in literature is not a sufficient indicator of or a measure of 'test mode effect' that may happen due to the very visual layout of the computer interface. Other unique learner characteristics should also be considered which may affect the students in a computer based assessment environment. Moreover since computerized delivery of the assessment contents provides test designers the flexibility to design innovative item types and presentation format it is argued that various presentation formats might affect students' performance in a computer based assessment environment. Therefore investigations have been carried out to optimize presentation formats that don't interfere with the students' individual characteristics and their performance. In other words, we compared between 'Presentation Formats" (specifically the interactive components of GUI's) of CBA's to investigate and answer the question whether Presentation Formats can be held responsible for variation in the users performance in an examination - where test content complexity as well as the intellectual capacity of the user is assumed constant throughout the various experimental conditions.

Five experiments were conducted based on the interface features like scrolling, navigational mechanism, customization and screen density (white space). The first experiment compared *scrolling* with a *without scrolling* CBA presentation format, the second experiment

Table 5.1

Consolidated Findings

CBA presentation format	Indian students (experts)	American students (experts)
Scrolling	Do not prefer	Prefer
Without scrolling	Prefer	Do not prefer
Vertical menu based navigation	Prefer	Prefer
Step based navigation	Prefer but vertical menu has a slight edge.	Prefer
Single question without customization	Prefer	Do not prefer
Question customization	Preferred when compared with time customization but did not prefer when compared with single question without customization	Prefer
Time customization	Do not prefer (performance affected)	Do not prefer (performance affected)
Low screen density (High white space)	Prefer	Prefer
High screen density (Low white space)	Prefer but Low screen density (High white space) has a slight edge.	Prefer

compared *vertical menu based navigation* with *step based navigation* presentation format, the third experiment compared *time customization* with *question customization* presentation format, the fourth experiment compared *single question without customization* with *question customization presentation* format and finally the fifth and the last experiment compared *low screen density* (high white space) with *high screen density* (low white space) CBA presentation formats. Students from India and USA participated in the experiments. All the participants were university students. First, they were screened for their computer attitude, computer anxiety and their level of acquaintance with CBA systems. Experts who had high computer attitude, low computer anxiety and had experience with CBA systems were allowed to participate in the main study. Ten aptitude questions were asked and twenty minutes were allotted to the participants for completing the test. Thereafter, they answered a series of questionnaire in order to capture their subjective experience during the CBA session. The entire study was delivered through the World Wide Web to the Indian and the American students. It is important to highlight here that

for the study that was carried out in USA, cohorts of students were asked to participate in a laboratory for the study. The Fulbright foundation sponsored the studies conducted in USA for which the researcher travelled down to USA to conduct the experiments in the American cultural settings.

The study employed a 2x2 factorial design. In each of the five experiments interface feature and cultural backgrounds of the students were the independent measures. Each of the two independent measures had two levels. Students' performance in terms of the test scores and time taken to complete the test was recorded. Their subjective experience during the assessment session was recorded in terms of subjective cognitive load, perceived system usability, perceived ease of use, perceived usefulness, perceived e-learner's satisfaction, perceived system quality, relative advantage and willingness to continue to use. A 2x2 factorial multivariate analysis was conducted. Results highlight that interface features do affect students subjective experience in a computer based assessment. It has also been observed that interface features can affect their performance. The results of the analysis have been argued from the instructional efficiency view. The consolidated findings of the experiments are discussed below.

5.2. Consolidated Findings

Table 5.1 (previous page) presents the consolidated findings of the five experiments. The findings from the experimental results vindicate the assumptions of the current research investigation. The results highlight that presentation format affects students. While the Indian students preferred a simple without scroll, without customization interface for CBA format, American students preferred with scroll and customizable format (questions). It has been found that interface feature like time customization can seriously affect the students' in terms of their scores and time required to complete the test. Therefore, the argument that presentation format can affect the students' in terms of their performance and their subjective evaluation of the assessment environment stands vindicated. It is important to highlight that though a between group study has been carried out but the groups were comparable in terms of gender distribution, mean age and percentage of marks obtained in the last qualifying exam. All the students were above average (from the perspective of Indian grading system) in terms of their marks they obtained in the last university exams. So groups were fairly homogeneous. Students who participated in the study had high compute attitude and low computer anxiety. All of them had previous experiences (more than once) with CBA systems. So the generalizations of the results are for the expert student groups who have fair experience with CBA environment. It can be argued here that if these expert student communities get affected with the format of the CBA presentation, the extent of effect of the presentation format on novice students (See Appendix 4, page 241) would be much more pronounced.

The results of the investigations reinforce the central argument of this thesis that special attention should be given while designing the presentation formats in a CBA system because if the students' performance gets affected due to it then, the goal of assessment would be lost and students would only try hard to get acquainted with the format of the presentation rather than forcing their intellectual abilities for solving the assessment problems.

5.3. Contribution of the thesis

The contribution of the research work has been divided into two broad categories. First, the findings of the research investigation would be discussed in the light of the published literature in "test mode" affect studies and then practical implication of the findings would be discussed for the user interface designers. A detailed argument elaborating the research findings from the perspective of the Culture variable is presented in Appendix 4, page 239.

5.3.1 Theoretical contribution

When the results of the five experiments are viewed from the perspective of the past research literature, it highlights the following:

1. Past research literature in scrolling did not find a significant difference in performance (test scores) between groups with and without scrolling (Hewson et al., 2007; Whitelock, 2009). In our study also, there was no significant effect of scrolling and without scrolling CBA presentation format on performance but the current study highlights that the subjective evaluation of these presentation formats are different for the two different student communities. While Indian students prefer without scrolling interface, American students on the other hand prefer with scroll interface. Now, given the case that the students' were experts, the effects of the interface may be pronounced for novice students. Moreover, from the instructional efficiency point of view it was found that without scroll is preferred by Indian students more because it invoked less cognitive load in comparison to the American students who felt without scrolling CBA presentation format invoked high cognitive load. Moreover, the study conducted by Hewson et al., 2007 and Whitelock, 2009 had not carried out a cross cultural investigation. In this thesis a cross cultural investigation was carried out and reported.

2. Navigation is important in a hypermedia based learning environment. It has been reported that navigational tools increase navigation efficiency, reduce the feeling of being lost and improve learning performance (Allison& Hammond, 1989; Dee-Lucas & Larkin, 1995; McDonald & Stevenson, 1999; Puntambekar et al., 2003). But it may also impede the cognitive flexibility of the learners by imposing a simplified structure over the existing knowledge structure (Gall & Hannafin, 1994) and as such may not compel the learners to fully interact with

the hypertext environment (Jonassen, 1986). The current investigation suggests that students in computer based assessment environment behave very differently than students in other components in hypermedia based learning environment. The results of the research investigation highlight that students in CBA do not get affected due to the navigational mechanisms namely step or vertical menu.

3. Past research investigations on item review and presentation found that performance are largely unaffected by whether examinees had the option to review, examinees prefer to have the option to review available (Leeson, 2006). But the current study highlighted that when customization features are used for facilitating item review and presentation students' performance get affected.

4. White space (Screen density): The published research literature has limited findings and few studies investigated this feature. Moreover, they studied the effect of white space in comprehension based test item layout but almost none reported about the effect of white space in MCQ (Multiple Choice Question) format and how graphical elements embedded in the GUI to aid the students in a test format and thus reducing the white space may affect the students. The current study investigates the issue of white space from this perspective and found that contrary to the effect of white space in comprehension based computer tests, in MCQ format based computer tests students' do not get affected due to the reduction or increase in white space.

5.3.2 Practical implications of the findings

The findings of the investigations highlight important design heuristics for GUI designers of CBA. While scrolling is preferred more by Americans than Indians, for designing a CBA system the designers should note that Indian students prefer single item without scroll presentation. Screen density does not seem to affect students in a CBA environment; hence visual aids in the GUI may assist students' performance. Indian students prefer simple, non-customizable interfaces whereas Americans prefer customized interfaces. These heuristics are helpful guides to the GUI designers in their efforts to minimize the effect of interface features on students' performance.

5.3.3 Summary of the contribution of the thesis

The summary of the contribution of the thesis are listed below:

1. The thesis provides answers to the theoretical contradictions of the research literature and highlights the effects of interface features on students' performance and other

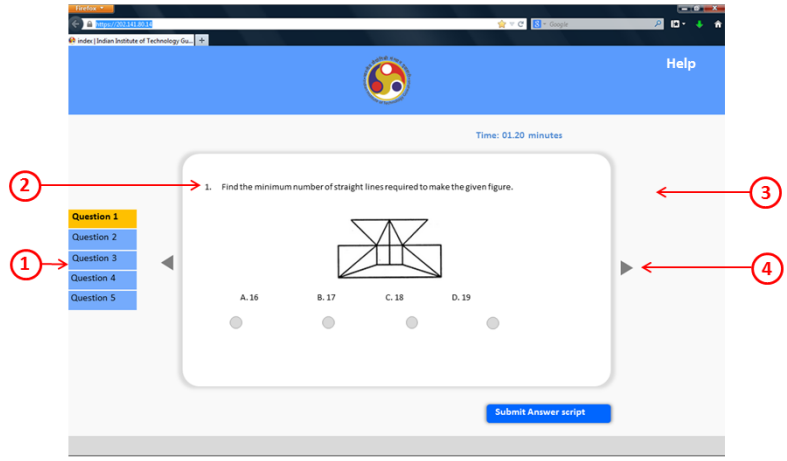


Figure 5.1. A CBA GUI design example for the Indian students based on the research findings of the thesis.

1 & 4 – Both step based and vertical menu based navigational mechanisms are used.

2. Single Item representation without scrolling

3. Maximum white space

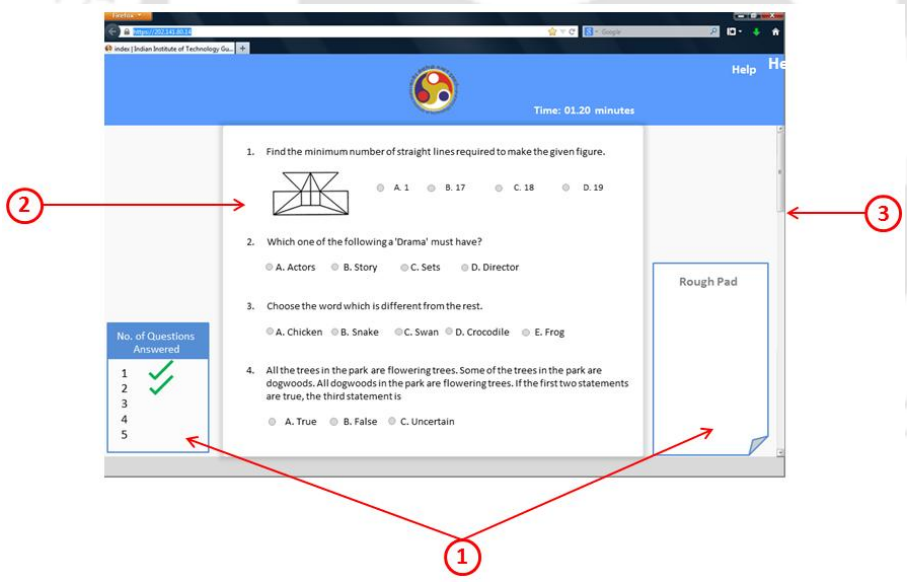


Figure 5.2. A CBA GUI design example for the US students based on the research findings of the thesis.

1 – Interactive elements for helping the students during assessment with reduced white space.

2 & 3 - Multiple Item representation with scrolling.

sensitive variables such as subjective cognitive load and perceived usability of the CBA system empirically.

2. It provides a verified frame of heuristics to Graphical user interface designers of computer based assessment systems to design the presentation format of test items both in

Indian as well as American cultural context. A CBA GUI design example incorporating the research findings has been illustrated in the figures 5.1 and 5.2.

3. It highlights the "presentation format effect" of various GUI features in a CBA system. Till now only "test mode effect" was known. This study for the very first time empirically investigated and documented the "presentation format effect" in addition.

4. This study provides the first cross cultural analysis of the effect of presentation format on test performance and the subjective evaluation of the CBA system by the students belonging to different cultural backgrounds (India & USA).

5.4. Inferences

This study highlights the effect of GUI features and cultural backgrounds on subjective cognitive load, usability features of a CBA system and students performance in CBA and argues that for reducing performance difference due to '*presentation effect*', individual characteristics alone should not only be considered but equal weightage should be given towards reducing the effect of Interface features on students.

5.5. Limitations and future scope

The study conducted in the Indian context allowed the participants to participate from their own choice of locations for assessing the internet but for the American context cohorts of students were allowed to participate from a specific location to assess the internet. The study should be replicated by capturing the learning styles of students so that effects can be verified. Moreover, the effects might be pronounced for students with higher test anxiety and lower computer skills. Future research should address this question (See Appendix 4, page 241). More effective way of capturing actual cognitive load invested by the students may provide a very definitive answer to the questions raised in the current thesis.

5.6. Chapter 5: Summary

The current research investigation highlighted that presentation format in CBA system can affect students' performance. They also affect the students' subjective evaluation of the CBA system. This may affect the CBA systems success and acceptance. Therefore, utmost care should be taken while designing the presentation format of these kinds of CBA systems because it has been shown that they can impede cognitive ability of the students. Moreover, User Interface Designers should consider the cultural factors of the students while designing these systems because these may interact with the interface features and affect the students capacity to perform during the assessment session.

Bibliography

A

Alderson, J. C. (2000). Technology in testing: The present and the future. *System, 28*(4), 593-603.

Association of Test Publishers. (2000).*Computer-based testing guidelines.* Washington, DC: Author.

Amelung, M., Krieger, K., & Rosner, D. (2011). E-Assessment as a Service. *IEEE Transactions on Learning Technologies, 4*(2), 162-174.

Anastasi, A., & Urbina, S. (1997). *Psychological testing (7th ed.)*. Englewood Cliffs, NJ: Prentice-Hall.

Ahern, S., & Beatty, J. (1979). Pupillary responses during information processing vary with Scholastic Aptitude Test scores. *Science*.

Allison, L., & Hammond, N. (1999). A learning support environment: The Hitch-Hiker's Guide. *In R. McAleese (Ed.), Hypertext: Theory into practice* (pp. 53-64). Exeter, England: Intellect Books.

Arbaugh, J. B. (2000). Virtual classroom characteristics and student satisfaction with internet-based MBA courses. *Journal of management education, 24*(1), 32-54.

B

Baek, S. G. (1994). Implications of cognitive psychology for educational testing. *Educational Psychology Review, 6*(4), 373-389.

Bahr, M. W., & Bahr, C. M. (1997). Educational assessment in the next millennium: Contributions of technology. *Preventing School Failure, 41*(2), 90-94.

Bunderson, C. V., Inouye, D. K., & Olsen, J. B. (1989). The four generations of computerized educational measurement. *In R. L. Linn (Ed.), Educational measurement (3rd ed.)* (pp. 367–407). NY: American Council on Education – Macmillan

Black, P., & Wiliam, D. (1998). Inside the Black Box: Raising Standards through Classroom Assessment. *Phi Delta Kappan, 80*(2),139-148.

Bresciani, M. J. (2002). The relationship between outcomes, measurement and decisions for continuous improvement. National Association for Student Personnel Administrators. *Net Results E-Zine*.

Bull, J. (1993). *The implementation and evaluation of computer-assisted assessment Computer-Assisted Assessment Centre.* http://caacentre.ac.uk/ - assessed on 19th January 2013.

Bugbee, A. C. (1996). The equivalence of paper-and-pencil and computer-based testing. *Journal of Research on Computing in Education, 28(*3), 282–299.

Birenbaum, M. (1996). Assessment 2000: Towards a pluralistic approach to assessment. *In M. Birenbaum & F. J. R. C. Dochy (Eds.), Alternatives in assessment of achievements, learning processes and prior knowledge* (pp. 3–29).Kluwer Academic Publications.

Bennett, R. E. (1998).*Reinventing assessment: Speculations on the future of large scale educational testing.* Princeton, NJ: Educational Testing Service, Policy Information Center.

Bodmann, S.M., & Robinson, D.H. (2004). Speed and performance differences among computer based and paper-pencil tests. *Journal of Educational Computing Research, 31*, 51–60.

Bunderson, C. V., Inouye, D. K., & Olsen, J. B. (1989). The four generations of computerized educational measurement. *In R. L. Linn (Ed.), Educational Measurement* (3rd ed, 367-408). New York: Macmillan.

Bennett, R. E., Braswell, J., Oranje, A., Sandene, B., Kaplan, B., & Yan, F. (2008). Does it matter if I take my mathematics test on computer? A second empirical study of mode effects in NAEP. The Journal of Technology, *Learning and Assessment, 6*(9).

Brunken, R., Plass, J. L., & Leutner, D. (2003). Direct measurement of cognitive load in multimedia learning. *Educational Psychologist, 38*(1), 53-61.

Bernard, M., Chaparro, B., & Thomasson, R. (2000). *Finding information on the web: Does the amount of whitespace really matter? Usability News, 2.1.* Retrieved January 2, 2013, from Software Usability Research Laboratory, Wichita State University: http://psychology.wichita.edu/surl/usabilitynews/42/text_length.htm.

Baddeley, A. (1992). Working memory. *Science, 255*(5044), 556-559.

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience, 4*(10), 829-839.

Barbeite, F. G., & Weiss, E. M. (2004). Computer self-efficacy and anxiety scales for an Internet sample: testing measurement equivalence of existing measures and development of new scales. *Computers in Human Behavior, 20*(1), 1-15.

Bickmore, T. W., & Picard, R. W. (2004, April). Towards caring machines. *In CHI'04 extended abstracts on Human factors in computing systems* (pp. 1489-1492). ACM.

Bray, J. H., & Maxwell, S. E. (1982). Analyzing and interpreting significant MANOVAs. *Review of Educational research, 52*(3), 340-367.

Brooke, J. (1996). SUS: A "quick and dirty" usability scale. *In: Jordan, P. W., Thomas, B.,Weerdmeester, B. A., McClelland (eds.) Usability Evaluation in Industry* pp. 189-194. Taylor & Francis.

C

Chatzopoulou, D. I., & Economides, A. A. (2010). Adaptive assessment of student's knowledge in programming courses. *Journal of Computer Assisted Learning, 26*(4), 258–269.

Clark, R. E. (1994). Media will never influence learning. *Educational technology research and development, 42*(2), 21-29.

Charman, D., & Elmes, A. (1998). *Computer based assessment: A guide to good practice*, 1. Plymouth: SEED Publications.

Cantillon, P., Irish, B., & Sales, D. (2004). Using computers for assessment in medicine. *British Medical Journal, 329*(7466), 606–609.

Clauser, B. E., Kane, M. T., Swanson, D. B. (2002). Validity issues for performance-based tests scored with computer-automated scoring systems. *Applied Measurement in Education, 15*(4), 413-432.

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), 593-602.

Charlton, J., & Hewson, C. (2007). Comparing Online and Offline Administration of Multiple Choice Question Assessments to Psychology Undergraduates: do assessment modality or computer attitudes influence performance?. *Psychology Learning & Teaching, 6*(1), 37-46.

Cassady, J. C., & Gridley, B. E. (2005). The effects of online formative and summative assessment on test anxiety and performance. *The Journal of Technology, Learning and Assessment, 4*(1).

Cheon, J., & Grant, M. (2009). Web-Based Assessment Tool for Communication and Active Listening Skill Development. *TechTrends: Linking Research and Practice to Improve Learning, 53*(6), 24-34.

Chiu, C., Hsu, M., Sun, S., Lin, T., Sun, P., 2005. Usability, quality, value and e-learning continuance decisions. *Computers & Education 45* (4), 399–416.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and instruction, 8*(4), 293-332.

Cheon, J., & Grant, M. M. (2009). Are pretty interfaces worth the time? The effects of user interface types on web-based instruction. *Journal of interactive learning research, 20*(1), 5-33.

D

Derbyshire, H. (1999). *The CASTLE toolkit.* http://www.le.ac.uk/castle/ – assessed on 23rd May 2013.

Drasgow, F., & Olsen-Buchanan, J. B. (1999). *Innovations in computerized assessment.* Mahwah, NJ: Erlbaum.

Delon, W., & Mclean, E. (1992). Information systems success: The quest for the dependent variable. *Information Systems Research, 3*(1), 60–95.

Drommi, A., Ulferts, G. W., & Shoemaker, D. (2001). Interface design: A focus on cognitive science. *In The Proceedings of ISECON 2001* (Vol. 18).

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, 319-340.

Dee-Lucas, D., & Larkin, J. H. (1995). Learning from electronic texts: Effects of interactive overviews for information access. *Cognitive and Instruction, 13*(3), 431-468.

DeAngelis, S. (2000). Equivalency of computer-based and paper-and-pencil testing. *Journal of allied health, 29*(3), 161-164.

Dillon, A., & Gabbard, R. (1998). Hypermedia as an educational technology: A review of the quantitative research literature on learner comprehension, control, and style. *Review of educational research, 68*(3), 322-349.

de Bruijn, D., de Mul, S., & van Oostendorp, H. (1992). The influence of screen size and text layout on the study of text. *Behaviour and Information Technology, 11*, 71–78.

Delon, W., & Mclean, E. (1992). Information systems success: The quest for the dependent variable. *Information Systems Research, 3*(1), 60–95.

Dhar, D., & Yammiyavar, P. (2012, July). Design Approach for E-learning Systems: Should it be User Centered or Learner Centered. *In Technology for Education (T4E), 2012 IEEE Fourth International Conference on* (pp. 239-240). IEEE.

Dhar, D., Adhikary, S., & Yammiyavar, P. (2012). An evaluation of the effect of navigational tools on cognitive load in a computer based test format. *In Intelligent Human Computer Interaction (IHCI), 2012 4th International Conference on* (pp. 1-6). IEEE.

Drommi, A., Ulferts, G. W., & Shoemaker, D. (2001). Interface design: A focus on cognitive science. *In The Proceedings of ISECON 2001* (Vol. 18).

E

Economides, A., & Roupas, C. (2007). Evaluating computer adaptive testing systems. *International Journal of Web-Based Learning and Teaching Technologies, 2*(1), 70–88.

Edgell, S. E., & Noon, S. M. (1984). Effect of violation of normality on the t test of the correlation coefficient. *Psychological Bulletin, 95*(3), 576.

F

Fluck, A., Pullen, D., & Harper, C. (2009).Case study of a computer based examination system. *Australasian Journal of Educational Technology, 25*(4), 509–523.

Fowler, S., & Stanwick, V. (2004). *Web application design handbook: Best practices for web-based software.* San Francisco, CA, USA: Morgan Kaufmann Publishers Inc..

Folk, L. C., March, J. Z., & Hurst, R. D. (2006). A comparison of linear, fixed-form computer-based testing versus traditional paper-and-pencil-format testing in veterinary medical education. *Journal of Veterinary Medical Education, 33*(3), 455-464.

Federico, P. A. (1989). Computer-based and paper-based measurement of recognition performance. Navy Personnel Research and Development Center Report NPRDC-TR-89-7.

G

Gvozdenko, E., & Chambers, D. (2007). Beyond test accuracy: Benefits of measuring response time in computerised testing. *Australasian Journal of Educational Technology, 23*(4), 542–558.

Gilliland, K., & Schlegel, R. E. (1993). *Readiness-to-perform testing: A critical analysis of the concept and current practices (FAA report no.FAA-AM-93-13).* Washington, DC: Department of Transportation, Federal Aviation Administration, Office of Aviation Medicine.

Gagné, R. M. (1985). *The conditions of learning and theory of instruction.* New York: Holt, Rinehart, and Winston.

Gikandi, J. W., Morrow, D., & Davis, N. E. (2011). Online formative assessment in higher education: a review of the literature. *Computers & Education, 57*(4), 2333–2351.

Greaud, V. A., & Green, B. F. (1986). Equivalence of conventional and computer presentation of speed tests. *Applied Psychological Measurement, 10*(1), 23-34.

Gall, J. E., & Hannafin, M. J. (1994). A framework for the study of hypertext. *Instructional Science, 22*(3), 207-232.

Garland, K.J., & Noyes, J.M. (2004). CRT monitors: Do they interfere with learning? *Behaviour & Information Technology, 23*, 43–52.

Goldberg, A. L., & Pedulla, J. J. (2002). Performance differences according to test mode and computer familiarity on a practice graduate record exam. *Educational and Psychological Measurement, 62*(6), 1053-1067.

Gohain, M. P. (2012, Dec 28). Record number of candidates for JEE 2013. *The Times of India.* Retrieved on February 21, 2013 from http://articles.timesofindia.indiatimes.com/2012-12-28/news/36035996_1_jee-apex-board-merit-list-delhi-and-other-cftis.

Gattiker, U. E., & Hlavka, A. (1992). Computer attitudes and learning performance: Issues for management education and training. *Journal of Organizational Behavior, 13*(1), 89-101.

Gefen, D., Karahanna, E., & Straub, D. W. (2003a). Inexperience and experience with online stores: the importance of TAM and trust. *Engineering Management, IEEE Transactions on, 50*(3), 307-321.

Gefen, D., Karahanna, E., & Straub, D. W. (2003). Trust and TAM in online shopping: an integrated model. *MIS quarterly*, 51-90.

Gefen, D., & Straub, D. W. (1997). Gender differences in the perception and use of e-mail: An extension to the technology acceptance model. *MIS quarterly*, 389-400.

H

Hsi, S., & Soloway, E. (1998). Learner-centered design: addressing, finally, the unique needs of learners. *In CHI 98 Conference Summary on Human Factors in Computing Systems* (pp. 211-212). ACM.

Hetter, R. D., & Sympson, J. B. (1997). Item exposure control in CAT-ASVAB. *In W. A. Sands, B. K. Waters, & J. R. McBride (Eds.), Computerized adaptive testing: From inquiry to operation* (pp. 141–144). Washington, DC: American Psychological Association.

Hewson, C. (2012). Can online course-based assessment methods be fair and equitable? Relationships between students' preferences and performance within online and offline assessments. *Journal of Computer Assisted Learning*, *28*(5), 488-498.

Honaker, L. M. (1988). The equivalency of computerized and conventional MMPI administration: a critical review. *Clinical Psychology Review, 8*(6), 561–577.

Haase, R. F., & Ellis, M. V. (1987). Multivariate analysis of variance. *Journal of Counseling Psychology, 34*(4), 404.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Human mental workload, 1*(3), 139-183.

Hsi, S., & Soloway, E. (1998, April). Learner-centered design: addressing, finally, the unique needs of learners. *In CHI 98 Conference Summary on Human Factors in Computing Systems* (pp. 211-212). ACM.

I

India: ICT led enterprise market is expected to be Rs 38,400 crore businesses by 2014!. (2012). Retrieved May 17, 2013 from http://telecomblogs.in/blog/2012/03/22/india-ict-led-enterprise-market-is-expected-to-be-rs-38400-crore-businesses-by-2014/

J

JISC, Effective Practice with e-Assessment; HEFCE, online, http://www.jisc.ac.uk/media/documents/themes/elearning/effpraceffpraceassess.pdf, 23rd May 2007.

Joosten-ten Brinke, D., van Bruggen, J., Hermans, H., Burgers, J., Giesbers, B., Koper,R., et al. (2007). Modeling assessment for re-use of traditional and new types of assessment. *Computers in Human Behavior, 23*(6), 2721–2741.

Jonassen, D. H. (1986). Hypertext principles for text and courseware design. *Educational Psychologist, 21*(4), 269-292.

Jeong, H. (2012). A comparative study of scores on computer-based tests and paper-based tests. *Behaviour & Information Technology*, (ahead-of-print), 1-13.

Joint Committee to Develop Standards for Educational and Psychological Testing of the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education (1985). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

K

Kane, R. L., & Kay, G. G. (1992). Computerized assessment in neuropsychology: *A review of tests and test batteries. Neuropsychology Review, 3*, 1–118.

Kaklauskas, A., Zavadskas, E. K., Pruskus, V., Vlasenko, A., Seniut, M., Kaklauskas, G.,et al. (2010). Biometric and intelligent self-assessment of student progress system. *Computers & Education, 55*(2), 821–833.

Kim, D. H., & Huynh, H. (2007). Comparability of computer and paper-and-pencil versions of algebra and biology assessments. *The Journal of Technology, Learning and Assessment, 6*(4).

Kim, D. H., & Huynh, H. (2008). Computer-based and paper-and-pencil administration mode effects on a state wide end-of-course English test. *Educational and Psychological Measurement, 68*(4), 554-570.

L

Lyell, R., & McNamara, S. (2000). Learning tool or potplant? Students' opinions of learning from a CAL program in a distance education context. *Australian Journal of Educational Technology, 16*(2), 126–146.

Lin, H., & Dwyer, F. (2006). The fingertip effects of computer-based assessment in education. *TechTrends, 50*(6), 27-31.

Luecht, R. M., & Sireci, S. (2011). *A Review of Models for Computer-Based Testing.* Retrieved on 21st January 2013 from https://research.collegeboard.org/sites/default/files/publications/2012/7/researchreport-2011-12-review-models-for-computer-based-testing.pdf

Leuthold, S., Schmutz, P., Bargas-Avila, J. A., Tuch, A. N., & Opwis, K. (2010). Vertical versus dynamic menus on the world wide web: Eye tracking study measuring the influence of menu design and task complexity on user performance and subjective preference. *Computers in human behavior, 27*(1), 459-472.

Leeson, H. V. (2006). The mode effect: A literature review of human and technological issues in computerized testing. *International Journal of Testing, 6*(1), 1-24.

Lee, H. K. (2004). A comparative study of ESL writers' performance in a paper-based and a computer-delivered writing test. *Assessing Writing, 9*(1), 4-26.

Lee, G., & Weerakoon, P. (2001). The role of computer-aided assessment in health professional education: A comparison of student performance in computer-based and paper-and-pen multiple-choice tests. *Med Teach, 23*(2), 152–157.

Lee, J. (1986). The effects of past computer experience on computer aptitude test performance. *Educational and Psychological Measurement, 46*,727–733.

Lynch, P. J., & Horton, S. (2002). *Web style guide: Basic design principles for creating web sites (2nd ed.)*. New Haven, CT: Yale University Press.

M

Mazzeo, J., & Harvey, A. L. (1988). *The equivalence of scores from automated and conventional educational and psychological tests: A review of the literature.* College Board Rep. No. 88-8. New York: College Entrance Examination Board.

Mead, A. D., & Drasgow, F. (1993). Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis. *Psychological Bulletin, 114*(3), 449–58.

McDonald, A. S. (2002). The impact of individual differences on the equivalence of computer-based and paper-and-pencil educational assessments. *Computers & Education, 39*(3), 299–312.

Marton, F., Booth, S., & Booth, S. A. (1997). *Learning and Awareness*. Pr. Routledge.

McMullin, J., Varnhagen, C. K., Heng, P., & Apedoe, X. (2002). Effects of surrounding information and line length on text comprehension from the web. *Canadian Journal of Learning and Technology, 28*, 19–29.

Mullet, K., & Sano, D. (1995). *Designing visual interfaces: Communication oriented techniques* .New York: Prentice Hall.

McDonald, S., & Stevenson, R. J. (1998). Navigation in hyperspace: An evaluation of the effects of navigational tools and subject matter expertise on browsing and information retrieval in hypertext. *Interacting with Computer, 10*, 129-142.

Mazzeo, J., Druesne, B., Raffeld, P. C., Checketts, K. T., & Muhlstein, A. (1991). *Comparability of computer and paper-and-pencil scores for two CLEP general examinations* College Board report No. 91-5.

Mason, B. J., Patry, M., & Berstein, D., J. (2001). An Examination of the equivalence between non-adaptive computer-based and traditional testing. *Journal of Educational Computing Research, 24*(1) 29–39.

Mousavi, S. Y., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of educational psychology, 87*(2), 319.

N

Nikolova, M. (2012).*Characteristics and Forms of the Electronic Assessment of the Knowledge*. http://conf.uni-ruse.bg/bg/docs/cp12/6.2/6.2-15.pdf assessed on 23rd May 2013.

Neil, P. M. (1996). The computerisation of the Army General Classification Test: A discussion paper. *Paper submitted for the 1996 E.F. Campbell Medal*. Canberra, Australia.

Nielson, F., Nielson, H. R., & Hankin, C. (1999). Principles of program analysis. Springer.

Novak, J. D., & Gowin, D.B. (1984). *Learning how to learn*. Cambridge University Press.

Noyes, J.M., Garland, K.J., and Robbins, E.L., 2004. Paper-based versus computer-based assessment: Is workload another test mode effect? *British Journal of Educational Technology, 35*, 111–113.

Noyes, J.M., & Garland, K.J. (2003). VDT versus paper-based text: Reply to Mayes, Sims and Koonce. *International Journal of Industrial Ergonomics, 31*, 411–423.

Noyes, J. M., & Garland, K. J. (2008). Computer-vs. paper-based tasks: are they equivalent?. *Ergonomics, 51*(9), 1352-1375.

O

Özalp-Yaman, S., & Çağıltay, N. E. (2010). Paper-based versus computer-based testing in engineering education. *In Education Engineering (EDUCON)*, pp. 1631-1637. IEEE.

Oviatt, S., Arthur, A., & Cohen, J. (2006). Quiet interfaces that help students think. *In Proceedings of the 19th annual ACM symposium on User interface software and technology* (pp. 191-200). ACM.

Oviatt, S. (2006). Human-centered design meets cognitive load theory: designing interfaces that help people think. *In Proceedings of the 14th annual ACM international conference on Multimedia* (pp. 871-880). ACM.

P

Parshall, C. G., Spray, J. A., Kalohn, J. C., & Davey, T. (2002). Considerations in Computer-Based Testing. *In Practical Considerations in Computer-Based Testing* (pp. 1-12). Springer New York.

Pellegrino, J.W., Chudowski, N., & Glaser, R. (Eds) (2001). *Knowing What Students Know*. Washington DC: National Academy of Sciences.

Prakash, L. S., & Saini, D. K. (2012). E-assessment for e-learning. *In Engineering Education: Innovative Practices and Future Trends (AICERA), 2012 IEEE International Conference on* (1-6). IEEE.

Perez, W. A., Masline, P. J., Ramsey, F. R., & Urban, K. E. (1987).*Unified Tri-Services Cognitive Performance Assessment Battery: Review and methodology (Technical Report AAMRL-TR-87)*. Wright-Patterson AFB, OH: Armstrong Aerospace Medical Research Laboratory.

Parshall, C. G., Spray, J. A., Kalohn, J. C., & Davey, T. (2002). *Practical considerations in computer-based testing*. New York: Springer.

Piccoli, G., Ahmad, R., & Ives, B. (2001). Web-based virtual learning environments: A research framework and a preliminary assessment of effectiveness in basic IT skills training. *Mis Quarterly*, 401-426.

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist, 38*, 63-71.

Puntambekar, S., Stylianou, A., & Hübscher, R. (2003). Improving navigation and learning in hypertext environments with navigable concept maps. *Human-Computer Interaction, 18*, 395-428.

Paas, F., & van Merriënboer, J. J. G. (1993).The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human Factors, 35*,737-743.

Poggio, J., Glasnapp, D. R., Yang, X., & Poggio, A. J. (2005). A comparative evaluation of score results from computerized and paper & pencil mathematics testing in a large scale state assessment program. The Journal of Technology, *Learning and Assessment, 3*(6).

Pedersen, P. E., & Nysveen, H. (2003, June). Usefulness and self-expressiveness: extending TAM to explain the adoption of a mobile parking service. *In 16th Electronic Commerce Conference*, Bled, Slovenia, June (pp. 9-11).

Poelmans, S., Wessa, P., Milis, K., Bloemen, E., & Doom, C. (2008). Usability and acceptance of e-learning in statistics education, based on the compendium platform. *In International Conference of Education, Research and Innovation (ICERI 2008)* (pp. 1-10).

Prümper, J., Frese, M., Zapf, D. & Brodbeck, F. C. (1991). Errors in computerized office work: differences between novice and expert users. *ACM SIGCHI Bulletin, 23*(2), 63-66.

R

Robitaille, D. F., Schmidt, W. H., Raizen, S., McKnight, C., Britton, E., & Nicol, C.(1993).*Curriculum frameworks for mathematics and science*. TIMSS Monograph, 1. Vancouver: Pacific Educational Press.

Ridgway, J., McCusker, S., & Pead, D. (2004). *Literature review of e-assessment.* Futurelab, Bristol.

Revuela, J., & Ponsoda, V. (1998). A comparison of item exposure control methods in computerized adaptive testing. *Journal of Educational Measurement, 35*, 311–327.

Rogers, E. M. (1983). *Diffusion of innovations*. New York: The Free Press.

Ricketts, C., & Wilks, S. J. (2002). Improving student performance through computer-based assessment: insights from recent research. *Assessment & Evaluation in Higher Education, 27*(5), 475-479.

Rao, Y. (2013, Sept. 26). It's paper & pencil for JEE-Advanced 2014. *The Times of India.* Retrieved on September 27, 2013 from http://timesofindia.indiatimes.com/city/mumbai/Its-paper-pencil-for-JEE-Advanced-2014/articleshow/22612043.cms.

S

Singh, R. S. (2012, Sept. 4). Personal computer sales grow 16% in 2011-12. *The Hindu Business Line*. Retrieved on March 21, 2013 from http://www.thehindubusinessline.com/industry-and-economy/info-tech/personal-computer-sales-grow-16-in-201112/article3858175.ece.

Stocking, M. L., & Lewis, C. (1995). *A new method for controlling item exposure in computerized adaptive testing (Research Report No. 95-25).* Princeton, NJ: Educational Testing Service.

Stocking, M. L., & Lewis, C. (1998). Controlling item exposure conditional on ability in computerized adaptive testing. *Journal of Educational and Behavioral Statistics, 23*, 57–75.

Saikkonen, R., Malmi, L., & Korhonen, A. (2001). Fully Automatic Assessment of Programming Exercises. *Proc. Sixth Ann. Conf. Innovation and Technology in Computer Science Education (ITiCSE '01)*, 133-136.

Schlegel, R. E., Hockey, G. R. J., & Gilliland, K. (2003). Performance tests. *In G. Wilson, & R. E. Schlegel (Eds.), Operator Functional State Assessment.* Neuilly Sur Seine, France: NATO Research and Technology Agency, Human Factors and Medicine Panel, Task Group 8.

Shingledecker, C. A. (1984). *A task battery for applied human performance assessment research. (Technical Report AFAMRL-TR-84-071).* Wright-Patterson AFB, OH: Air Force Aerospace Medical Research Laboratory.

Scalise, K., & Gifford, B. (2006). Computer-based assessment in e-learning: A framework for constructing ''Intermediate Constraint'' questions and tasks for technology platforms. Journal of Technology, *Learning, and Assessment, 4*(6).

Segall, D. O., & Moreno, K. E. (1999). Development of the computerized adaptive testing version of the Armed Services Vocational Aptitude Battery. *In F. Drasgow & J. B. Olsen-Buchanan (Eds.), Innovations in computerized assessment* (pp. 35–65). Mahwah, NJ: Erlbaum.

Sympson J. B., & Hetter, R. D. (1985). Controlling item exposure rates in computerized adaptive tests. *Paper presented at the Annual Conference of the Military Testing Association*, San Diego, CA.

Stoney, S., & Wild, M. (1998). Motivation and interface design: Maximizing learning opportunities. *Journal of Computer Assisted Learning, 14*, 40-50.

Seddon, P. B. (1997). A respecification and extension of the DeLone and McLean model of IS success. *Information systems research, 8*(3), 240-253.

Spool, J., Scanlon, T., Schroeder, W., Snyder, C., & De Angelo, T. (1999). *Web site usability: A designer's guide.* San Francisco: Kaufmann.

Schaeffer, G. A., Reese, C. M., Steffen, M., McKinley, R. L., & Mills, C. N. (1993). *Field test of a computer-based GRE general test.* ETS Research Report #93-07.

Sutton, R. E. (1997). Equity and high stakes testing: implications for computerized testing. *Equity and Excellence in Education, 30*(1), 5–15.

Saadé, R. G., & Kira, D. (2009). Computer anxiety in e-learning: The effect of computer self-efficacy. *Journal of Information Technology Education, 8*(177-191).

Selwyn, N. (1997). Students' attitudes toward computers: Validation of a computer attitude scale for 16–19 education. *Computers & Education, 28*(1), 35-41.

Sun, P. C., Tsai, R. J., Finger, G., Chen, Y. Y., & Yeh, D. (2008). What drives a successful e-Learning? An empirical investigation of the critical factors influencing learner satisfaction. *Computers & Education, 50*(4), 1183-1202.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive science, 12*(2), 257-285.

T

Tseng, H., Macleod, H. A., & Wright, P. (1997). Computer anxiety and measurement of mood change. *Computers in Human Behavior, 13*(3), 305–316.

Turner, G., & Gibbs, G. (2010). Are assessment environments gendered? An analysis of the learning responses of male and female students to different assessment environments. *Assessment & Evaluation in Higher Education, 35*(6), 687–698.

Taras, M. (2005). Assessment–summative and formative–some theoretical reflections. *British Journal of Educational Studies, 53*(4), 466–478.

Thelwall, M. I. K. E. (2000). Computer-based assessment: a versatile educational tool. *Computers & Education, 34*(1), 37-49.

Tsai, T. H., & Shin, C. D. (2012). A Score Comparability Study for the NBDHE Paper–Pencil Versus Computer Versions. *Evaluation & the Health Professions, 36*(2), 228-239.

Tseng, H. M., Tiplady, B., Macleod, H. A., & Wright, P. (1998). Computer anxiety: A comparison of pen-based personal digital assistants, conventional computer and paper assessment of mood and performance. *British Journal of Psychology, 89*(4), 599-610.

Tindall-Ford, S., Chandler, P., & Sweller, J. (1997). When two sensory modes are better than one. *Journal of experimental psychology: Applied, 3*(4), 257.

U

Underwood, J. D. M., & Brown, J. (1997). *Integrated learning systems: potential into practice*. London: Heineman.

V

Van de Vijver, F. J. R., & Harsveld, M. (1994). The incomplete equivalence of the paper-and-pencil and computerized versions of the General Aptitude Test Battery. *Journal of Applied Psychology, 79*(6), 852–859.

van Nes, F. (1986). Space, colour and typography of visual display terminals. *Behaviour and Information Technology, 5*, 99–118.

Vincent, A., Craik, F.I.M., & Furedy, J.J. (1996). Relations among memory performance, mental workload and cardiovascular responses. *International Journal of Psychophysiology, 23*, 181–198.

van De Velde, C., & von Gru¨nau, M. (2003). Tracking eye movements while reading: Printing press versus the cathode ray tube. *In: Proceeding of the 26th European conference on visual perception, ECVP 2003*. Paris, France: ECVP, 107.

Van Merrienboer, J. J., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational psychology review, 17*(2), 147-177.

W

Whiting, J. (1985).The use of a computer tutorial as a replacement for human tuition in a mastery learning strategy. *Computers and Education*, *9*(2), 101-109.

Whitelock, D. (2009). Editorial: e-assessment: developing new dialogues for the digital age. British *Journal of Educational Technology, 40*(2), 199-202.

Wa¨stlund, E., et al. (2005). Effects of VDT and paper presentation on consumption and production of information: Psychological and physiological factors. *Computers in Human Behavior, 21*, 377–394.

Wallace, P. E., & Clariana, R. B. (2000). Achievement predictors for a computer-applications module delivered via the world-wide web. *Journal of Information Systems Education, 11*(1)13–18. [http://gise.org/JISE/Vol11/v11n1-2p13-18.pdf]

Watson, B. (2001). Key factors affecting conceptual gains from CAL. *British Journal of Educational Technology, 32*(5) 587–593.

Wickens, C. D., (2002). Multiple resources and performance prediction. *Theoretical issues in ergonomics science, 3*(2), 159-177.

Wickens, C. D., Sandry, D. L., & Vidulich, M. (1983). Compatibility and resource competition between modalities of input, central processing, and output. Human Factors: *The Journal of the Human Factors and Ergonomics Society, 25*(2), 227-248.

Y

Yammiyavar, P., & Dhar, D. (2011). Conceptualizing GUIs—A Case Study of an Online Aptitude Testing System. *Research into Design-Supporting Sustainable Product Development (ICoRD'11)*, 115-124.

Z

Zakrzewski, S., & Bull, J. (1998).The mass implementation and evaluation of computer-based assessments. *Assessment and Evaluation in Higher Education, 23*(2), 141-152.

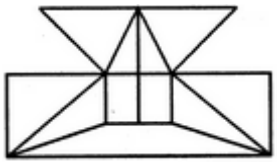
Ziefle, M., 1998. Effects of display resolution on visual performance. *Human Factors*, *40*(4), 554–568.

Appendix-1

CBA Test Items

Experiment: With Scroll vs. Without Scroll.

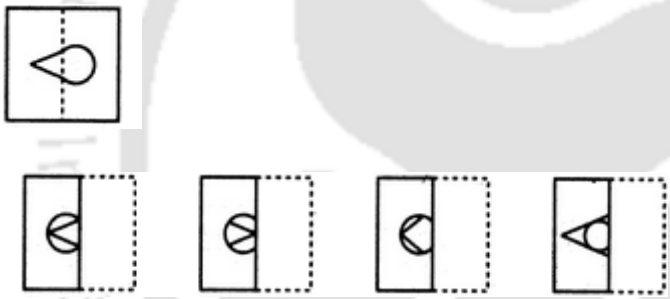
1. Find the minimum number of straight lines required to make the given figure.



A. 16 B. 17 C. 18 D. 19

Answer: B.

2. Find out from amongst the four alternatives as to how the pattern would appear when the transparent sheet is folded at the dotted line.



(1) (2) (3) (4)

Answer: A.

3. Select a suitable figure from the Answer Figures that would replace the question mark (?).

Problem Figures: Answer Figures:

(A) (B) (C) (D) (1) (2) (3) (4) (5)

A. 1 B. 2 C. 3 D. 4 E. 5

Answer: C.

4. Which one of the following a 'Drama' must have?

A. Actors B. Story C. Sets D. Director

Answer: B.

5. Choose the word which is different from the rest.

A. Chicken B. Snake C. SwanD. Crocodile E. Frog

Answer: A.

6. All the trees in the park are flowering trees. Some of the trees in the park are dogwoods.

All dogwoods in the park are flowering trees. If the first two statements are true, the third statement is

A. True B. False C. Uncertain

Answer: A.

7. Which word does NOT belong with the others?

A. Inch B. Ounce C. Centimetre D. Yard

Answer: B.

8. **Statements:** The question below has two statements I and II. These statements may be either independent causes or may be effects of independent causes or a common cause. One of these statements may be the effect of the other statements. Read both the statements and decide which of the following answer choice correctly depicts the relationship between these two statements.

I. The meteorological Department has issued a statement mentioning deficient rainfall during monsoon in many parts of the country.

II. The Government has lowered the revised estimated GDP growth from the level of earlier estimates.

A. Statement I is the cause and statement II is its effect.

B. Statement II is the cause and statement I is its effect.

C. Both the statements I and II are independent causes.

D. Both the statements I and II are effects of independent causes.

E. Both the statements I and II are effects of some common cause.

Answer: D.

9. Clorida is taller than Ivory. Emily is taller than Lovely. Lovely is taller than Enamol. To determine who among them is the tallest, which of the following further information, if any, is required?

A. Clorida is taller than Enamol and Lovely.

B. No further information is needed.

C. Emily is taller Ivory.

D. Clorida is taller than Lovely.

E. Enamol is taller than Clorida.

Answer: E.

10. Quber, a specialty bistro, is open for people every Monday to Saturday but it is closed on Sundays. The bistro gives out only lunch on Mondays, Tuesdays as well as Thursdays, while it gives out only supper on Wednesdays, Fridays, as well as Saturdays. On days that Quber is open for people, floors are cleaned and trees are watered according to the following guidelines:

I. Trees are watered two days every week, but not at all on repeated days and never on the same day are the floors cleaned

II. Floors are cleaned three days every week, but never on repeated days and never on the same day are that trees watered

If dinner is given out on a day that trees are watered, which of the following sentence should be true?

A. Floors are cleaned on Thursday

B. Trees are watered on Saturday

C. Floors are cleaned on Wednesday

D. Trees are watered on Tuesday

E. Trees are watered on Wednesday

Answer: A.

Question References:

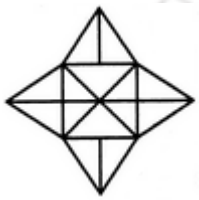
www.indiabix.com

www.coolinterview.com

www.bestsamplequestions.com

Experiment: Vertical Menu based Navigation vs. Step based Navigation

1. Find the number of triangles in the given figure.



A. 18 B. 20 C. 28 D. 34

Answer: C.

2. David ranks seventh from the top and 28th from the bottom in a class. How many students are there in the class?

A. 36 B. 35 C. 34 D. Cannot be determined E. None of the above

Answer: C.

3. Four of the following five are similar in a definite way and so form a group. Which one of them does not belong to the group?

A. Umbrella B. Gloves C. Shirt D. Shoes E. Cap

Answer: A.

4. Joseph, Jackie and Ninda are the three mountaineers. Joseph is Jackie's brother. Jackie is Ninda's brother. Ninda is not Joseph's brother. Therefore, Ninda is Joseph's sister.

A. Should be true B. May be false C. Should be false D. Cannot be determined

Answer: C.

5. A train always has

A. Rails B. Driver C. Engine D. Guard.

Answer: C.

6. Statements: (Find which one is the cause and which one is the effect)

1. The staff of Airport Authorities called off the strike they were observing in protest against privatization.

2. The staff of Airport Authorities went on strike anticipating a threat to their jobs.

A. Statement I is the cause and statement II is its effect.

B. Statement II is the cause and statement I is its effect.

C. Both the statements I and II are independent causes.

D. Both the statements I and II are effects of independent causes.

E. Both the statements I and II are effects of some common cause.

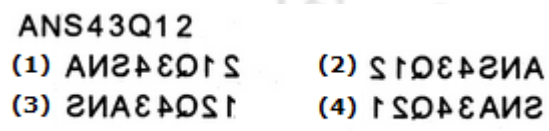
Answer: D.

7. Which word does NOT belong with the others?

A. Tulip B. Rose C. Bud D. Daisy

Answer: C.

8. Choose the alternative which is closely resembles the mirror image of the given combination.



A. 1 B. 2 C.3 D. 4

Answer: B.

9. The question has an underlined word followed by four answer choices. You will choose the word that is a necessary part of the underlined word.

School

A. Student B. Report card C. Test D. Learning

Answer: A.

10. Find out the relationship

Window is to pane as book is to

A. Novel B. Glass C. Cover D. Page

Answer: D.

Question Reference:

www.indiabix.com

www.bestsamplequestions.com

Experiment: Question Customization & Single Question without Customization

1. Find the number of triangles in the given figure.

A. 8 B. 10 C. 12 D. 14

Answer: D.

2. In the following question, five groups of letters have been given, out of which four are alike in some way and one is different. Choose the odd one out.

A. Treat B. Great C. Tears D. Rates E. Heard

Answer: E.

3. P Y 4 # A Q © K B 8 D 5 * E % 2 S ? 9 R U $ F @ 6 H M 3 W (X)

If all the symbols are dropped from the sequence, then which element will be in the centre in the new sequence?

A. 5 B. E C. 2 D. S E. There is no centre element

Answer: E.

4. In the following question, four groups of letters have been given, out of which four are alike in some way and one is different. Choose the odd one out.

A. Curtain B. Saree C. Shirt D. Petticoat

Answer: A.

5. A child should not be----as being either very shy or over aggressive.

A. Categorized B. Instructed C. Intoned D. Unfocused E. abstained

Answer: A.

6. Starting in 1985, all drivers had to wear helmets if they have to drive a two wheeler. From which of the following can this statement be properly inferred?

A. Some drivers may have worn helmets before 1985, but all two wheeler drivers were required to wear them beginning in 1985.

B. No drivers had to wear helmets before 1985.

C. Two wheelers drivers were the first to be required to wear helmets.

D. Two wheelers drivers had to wear helmets prior to 1985.

E. None of the above.

Answer: A.

7. **Statements:**

I. The staff of Airport Authorities called off the strike they were observing in protest against privatization.

II. The staff of Airport Authorities went on strike anticipating a threat to their jobs.

A. Statement I is the cause and statement II is its effect.

B. Statement II is the cause and statement I is its effect.

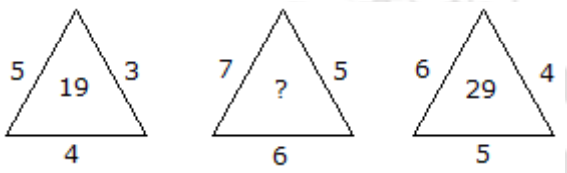
C. Both the statements I and II are independent causes.

D. Both the statements I and II are effects of independent causes.

E. Both the statements I and II are effects of some common cause.

Answer: D.

8. Which one will replace the question mark?



A. 18 B. 12 C.9 D.6

Answer: C.

9. Five girls are sitting on a bench to be photographed. Seema is to the left of Rani and to the right of Bindu. Mary is to the right of Rani. Reeta is between Rani and Mary.

Who is sitting immediate right to Reeta ?

A. Bindu B. Rani C. Mary D. Seema

Answer: C.

10. Tanya is older than Eric.

Cliff is older than Tanya.

Eric is older than Cliff.

If the first two statements are true, the third statement is

A. True B. False C. Uncertain D. Cannot predict

Answer: B.

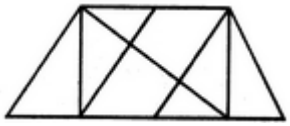
Question Reference:

www.indiabix.com

www.bestsamplequestions.com

Experiment: Question Customization vs. Time Customization

1. Find the number of triangles in the given figure.



A. 12 B. 18 C. 22 D. 26

Answer: B.

2. In the following question, five groups of letters have been given, out of which four are alike in some way and one is different. Choose the odd one out.

A. Great B. Treat C. Tears D. Table F. Later

Answer: D.

3. P Y 4 # A Q © K B 8 D 5 * E % 2 S ? 9 R U $ F @ 6 H M 3 W (X)

#8 is related to? W in the same way as BE is related to?

A. S? B. 6M C. W(x) D. $3 E. None of the above

Answer: E.

4. In the following question, four groups of letters have been given, out of which four are alike in some way and one is different. Choose the odd one out.

A. Joy B. Love C. Hate D. Toy

Answer: D.

5. In his general----manner, he had covered himself against this type of loss.

A. Thoughtful B. Fortunate C. Uncaring D. Circumspect E. Scathing

Answer: D.

6. In 1970, the average child visited the dentist once a year. By 1990 the number of visits increased to two. Today, the average child visits the dentist three times a year. Each of the following, if true, could explain this trend except.

A. Better home care of teeth has reduced the number of cavities

B. Dentist fee have declined over the period

C. Parents are more aware of the importance of dental care

D. Dental care has become less painful

E. None of the above

Answer: C.

7. **Statements:**

1. A huge truck overturned on the middle of the road last night.

2. The police had cordoned of entire area in the locality this morning for half of the day.

A. Statement I is the cause and statement II is its effect.

B. Statement II is the cause and statement I is its effect.

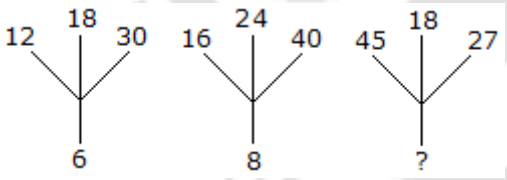
C. Both the statements I and II are independent causes.

D. Both the statements I and II are effects of independent causes.

E. Both the statements I and II are effects of some common cause.

Answer: A.

8. Which one will replace the question mark?



A. 25 B. 37 C. 41 D. 47

Answer: C.

9. Five girls are sitting on a bench to be photographed. Seema is to the left of Rani and to the right of Bindu. Mary is to the right of Rani. Reeta is between Rani and Mary.

Who is second from the right?

A. Mary B. Rani C. Reeta D. Bindu

Answer: C.

10. Tanya is older than Eric.

Cliff is older than Tanya.

Eric is older than Cliff.

If the first two statements are true, the third statement is

A. True B. False C. Uncertain D. Cannot predict

Answer: B.

Question Reference:

www.indiabix.com

www.bestsamplequestions.com

Experiment: High white space vs. Low white space

1. Pick out the most effective word(s) from the given words to fill in the blank to make the sentence meaningfully complete.

Fate smiles ..... those who untiringly grapple with stark realities of life.

A. With B. Over C. On D. Round

Answer: C.

2. The question has an underlined word followed by four answer choices. You will choose the word that is a necessary part of the underlined word.

Harvest

A. Autumn B. Stockpile C. Tractor D. Crop

Answer: D.

3. All the trees in the park are flowering trees.

Some of the trees in the park are dogwoods.

All dogwoods in the park are flowering trees.

If the first two statements are true, the third statement is

A. True B. False C. Uncertain D. Cannot Answer

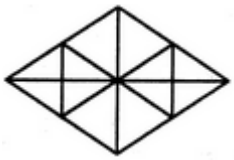
Answer: A.

4. Odometer is to mileage as compass is to

A. Speed B. Hiking C. Needle D. Direction

Answer: D.

5. Find the number of triangles in the given figure.



A. 16 B. 22 C. 28 D. 32

Answer: C.

6. The question has an underlined word followed by four answer choices. You will choose the word that is a necessary part of the underlined word

Pain

A. Cut B. Burn C. Nuisance D. Hurt

Answer: D.

7. A, B, C, D and E are sitting on a bench. A is sitting next to B, C is sitting next to D, D is not sitting with E who is on the left end of the bench. C is on the second position from the right. A is to the right of B and E. A and C are sitting together. In which position A is sitting?

A. Between B and D B. Between B and C C. Between E and D D. Between C and E

Answer: B.

8. Which one of the following is always found in 'Wonder'?

A. Crowd B. Lumber C. Astonishment D. Rustic

Answer: C.

9. If D is the brother of B, how B is related to C? To answer this question which of the statements is/are necessary?

1. The son of D is the grandson of C.

2. B is the sister of D.

A. Only 1 B. Only 2 C. Either 1 or 2 D. 1 and 2 both are required

Answer: D.

10. A pineapple costs Rs. 7 each. A watermelon costs Rs. 5 each. X spends Rs. 38 on these fruits. The number of pineapples purchased is

A. 2 B. 3 C.4 D. Data inadequate

Answer: C.

Question Reference:

www.indiabix.com

www.education.com

www.coolinterview.com

www.m4maths.com

Appendix- 2A

Screening Questionnaire*

1. Did you ever take Computer based tests previously?

☐ Yes No ☐

(If the answer is yes, the participant receives the next question)

2. Choose the correct option from the options given below:

i. I have given Computer based tests only once. ☐

ii. I have given Computer based tests more than once. ☐

3. Learner's attitude towards computers questionnaire

1. Is very difficult. (R)
2. Is very complicated. (R)
3. Requires technical ability. (R)
4. let me feel psychological stress very greatly. (R)
5. Can be done only if one knows a programming language such as basic. (R)
6. Is only advisable for people with a lot of patience. (R)
7. Makes a person more productive at his/her job. (R)
8. Is for young people only. (R)

(Likert's scale 1, strongly disagree; 7, strongly agree)

(R) Reverse Coded.

3. Learner computer anxiety questionnaire

1. Working with a computer would make me very nervous.
2. I get a sinking feeling when I think of trying to use a computer.
3. Computers make me feel uncomfortable.
4. Computers make me feel uneasy and confused.

(Likert's scale 1, strongly disagree; 7, strongly agree)

* Questionnaires were delivered to the participants through the web (internet).

Appendix -2B

Pre-test Questionnaire*

The following Items were asked in the pre-test questionnaire.

1. Age

2. Gender

3. Country of Permanent Residence

4. Ethnicity[1]

5. Degree

6. Major

7. Percentage of Marks obtained in the last qualifying exam. (E.g. percentage of marks obtained in the last semester).

[1]This item was primarily given for USA participants. The options were:

i. White.

ii. Black or African American.

ii. American Indian or Alaska Native.

iii. Asian.

iv. Native Hawaiian.

v. Some other race. Mention……..

*** Questionnaires were delivered to the participants through the web (internet).**

Appendix- 2C

Post-test Questionnaire

The entire post-test questionnaire was delivered to the participants through the web.

System Usability Scale

1. I think that I would like to use this system frequently.

2. I found the system unnecessarily complex.

3. I thought the system was easy to use.

4. I think that I would need the support of a technical person to be able to use this system.

5. I found the various functions in this system were well integrated.

6. I thought there was too much inconsistency in this system.

7. I would imagine that most people would learn to use this system very quickly.

8. I found the system very cumbersome to use.

9. I felt very confident using the system.

10. I needed to learn a lot of things before I could get going with this system.

(Likert scale 1, Strongly disagree; 5, Strongly agree)

Perceived ease of use scale

1. It was simple to use this online exam*.

2. I feel comfortable using this online exam*.

3. It was easy to learn to use this online exam*.

4. I believe I became productive quickly using this online exam*.

(Likert scale 1, Strongly disagree; 5, Strongly agree)

*The word "website" in the original item was changed to "online exam" for the current study.

Perceived usefulness scale

1. I can effectively complete my task[1] using this online exam[2] system.

2. I am able to complete my task[1] quickly using this online exam[2] system.

3. I am able to efficiently complete my task[1] using this online exam[2] system.

(Likert scale 1, Strongly disagree; 5, Strongly agree)

[1]The word "work" in the original item was changed to "task" for the current study.

[2]The word "website" in the original item was changed to "online exam" for the current study.

System quality scale

1. Whenever I make a mistake using the website, I recover easily and quickly.

2. The information provided with this online exam to complete the task is clear[1].

3. The interface of this online exam[2] is pleasant.

4. I like using the interface of this online exam[2].

5. This online exam[2] has all the functions and capabilities I expect it to have.

6. The online exam[2] contains sufficient hyperlinks to navigate.

7. I found scrolling frustrating when using the online exam[2]. (R)

8. It is easy to move from one page to another.

9. Overall, I found that the online exam[2] was able to quickly perform interactive tasks[3].

10. Overall, the online exam[2] reacts quickly and I don't have to wait too long to go to a new page[4].

11. The software is stable and doesn't crash regularly.

(Likert scale 1, Strongly disagree; 5, Strongly agree)

[1]The original item was slightly modified in this study. The sentence "such as online help, online messages, and other documentation" was removed after the word information from the original item.

[2]The word "website" from the original item was changed to "online exam" for the current study.

[3]The word "statistical computations" from the original item was changed to "interactive tasks" for the current study.

[4]The word "download information" was removed from the original item.

(R) Reverse Coded.

Relative advantage scale

1. Taking Exam, through this online exam system is more effective than taking Exam in traditional pen and paper based format[1].

2. This online exam system is better than other online exam systems I have had so far[2].

(Likert scale 1, Strongly disagree; 5, Strongly agree)

[1]The original item was slightly modified in this study. The original scale compared between learning statistics through a website to the traditional handbook.

[2] The original item was slightly modified in this study. The original scale compared between learning statistics through a website to other statistical courses.

WCU Scale

Please rate your willingness to continue working with the software in future Examinations

(1-10 rating scale; 1, Lowest; 10, Highest)

Perceived e-learner satisfaction scale

1. I am satisfied to give this examination* via the Internet

2. If I had an opportunity to take another examination* via the Internet, I would gladly do so.

3. My choice to take this examination* via the Internet was a wise one.

4. I was very satisfied with the examination*

5. I feel that this exam* served my needs well.

6. I will take as many examinations* via Internet as I can.

7. I was disappointed with the way the examination* worked out. (R)

8. If I had it to do over, I would not take this examination* via the Internet. (R)

9. Conducting the examination* via the Internet made it more difficult than other examination* I have given. (R)

(Likert scale 1, Strongly disagree; 7, Strongly agree)

*The word course of the original course was changed to examination and exam here.

(R) Reverse Coded.

NASA Task Load Index

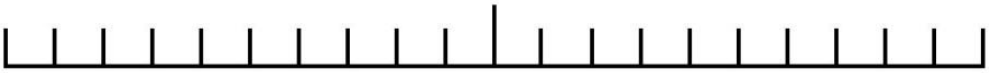
Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.

Name	Task	Date

Mental Demand — How mentally demanding was the task?

Very Low — Very High

Physical Demand — How physically demanding was the task?

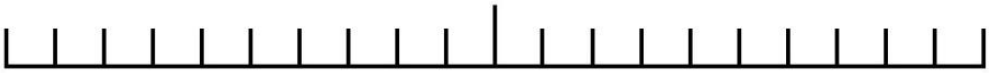


Very Low — Very High

Temporal Demand — How hurried or rushed was the pace of the task?

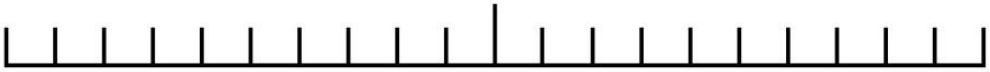
Very Low — Very High

Performance — How successful were you in accomplishing what you were asked to do?



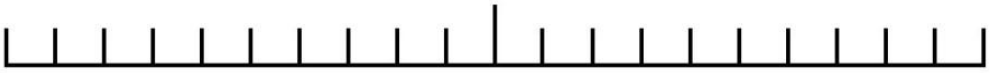
Perfect — Failure

Effort — How hard did you have to work to accomplish your level of performance?



Very Low — Very High

Frustration — How insecure, discouraged, irritated, stressed, and annoyed wereyou?



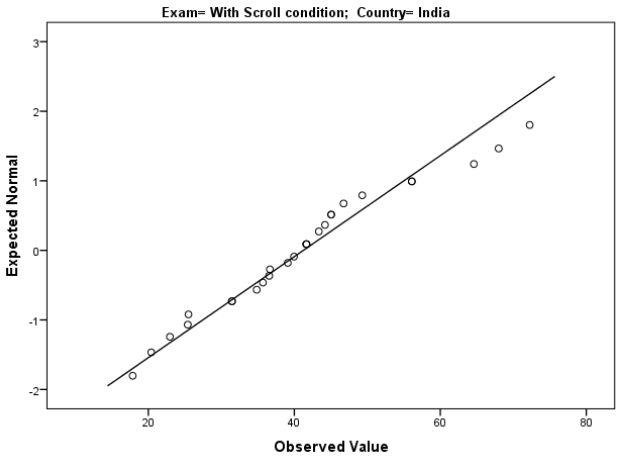
Very Low — Very High

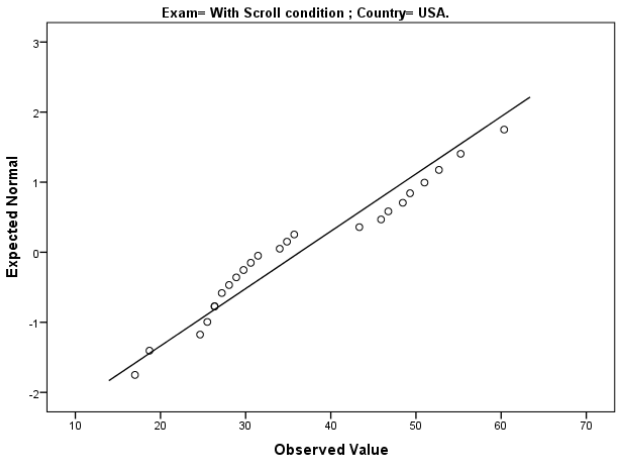
Appendix- 3A

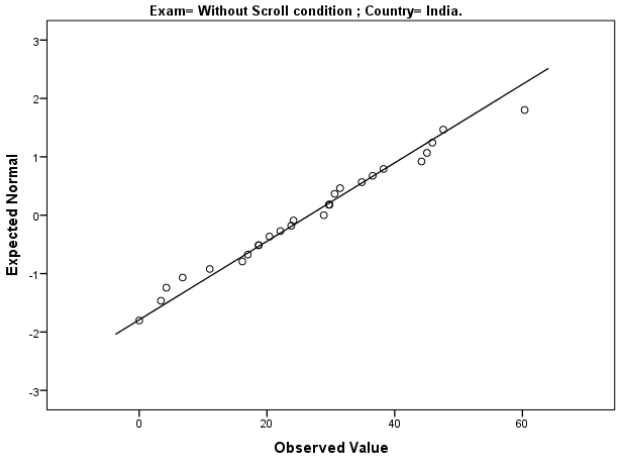
Experiment 1: Scrolling vs. Without Scrolling

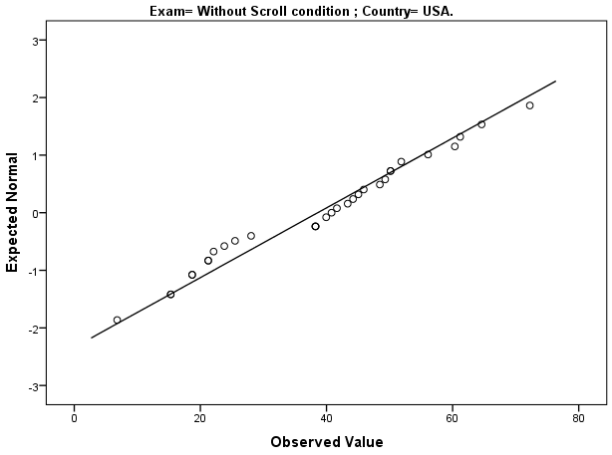
Figure 3A.1 Normal Q-Q plots

NASA TLX

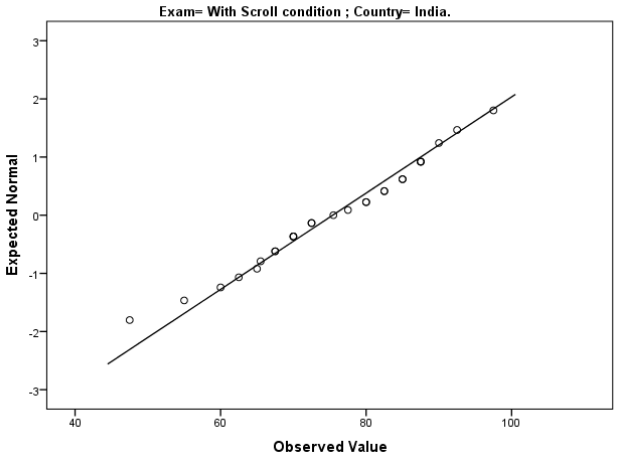


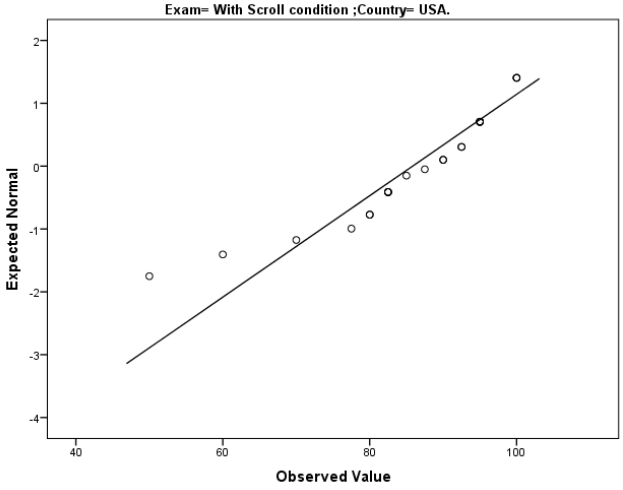


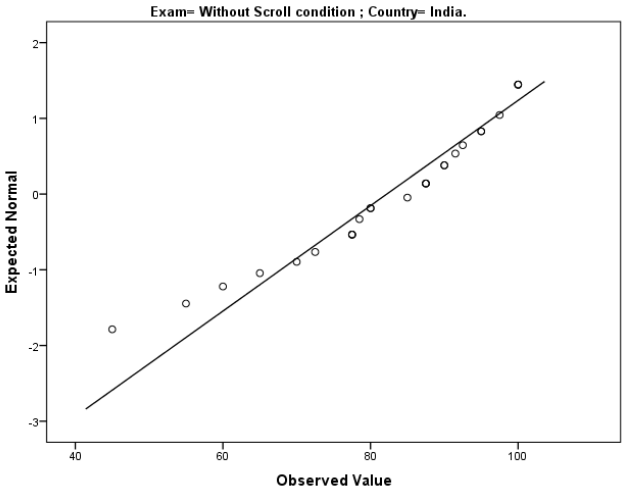


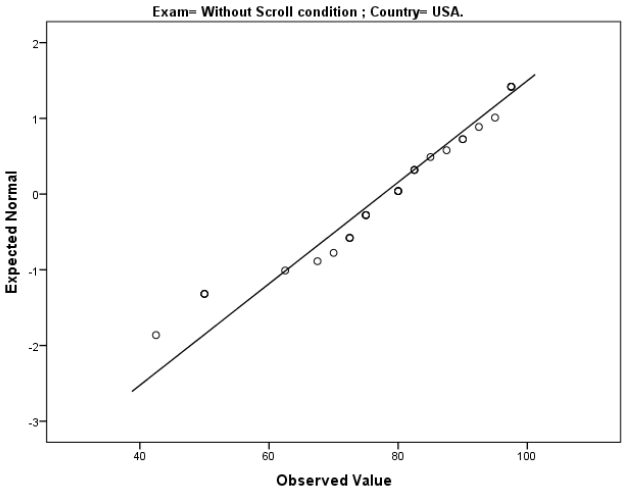


Perceived System Usability

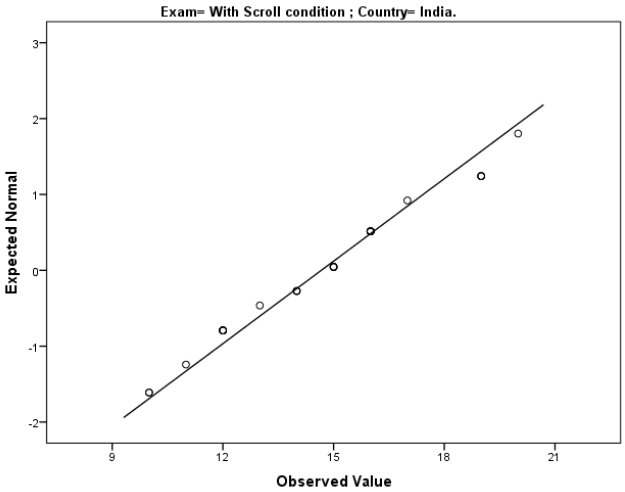


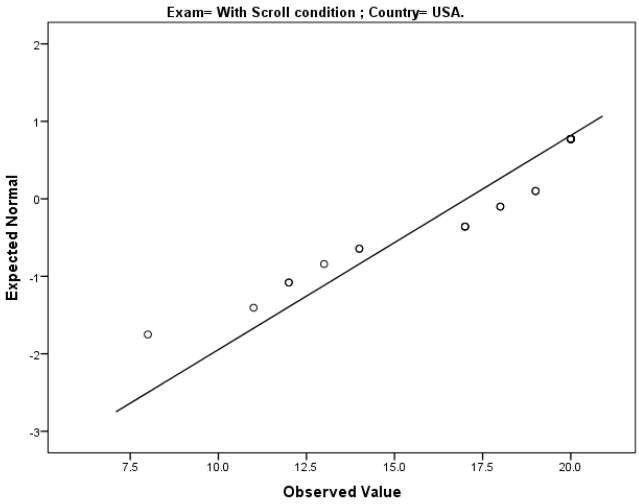


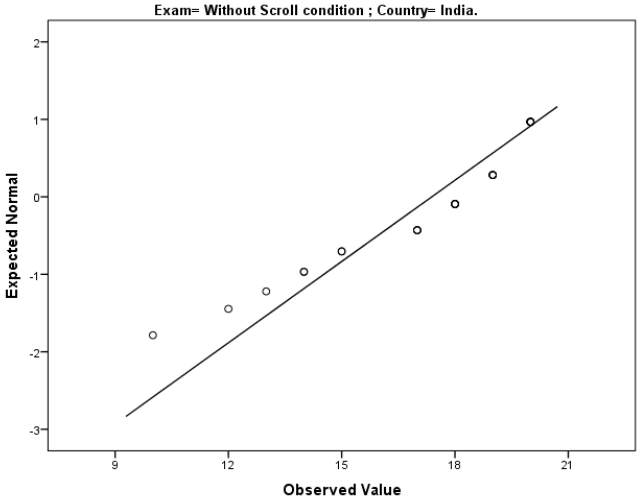


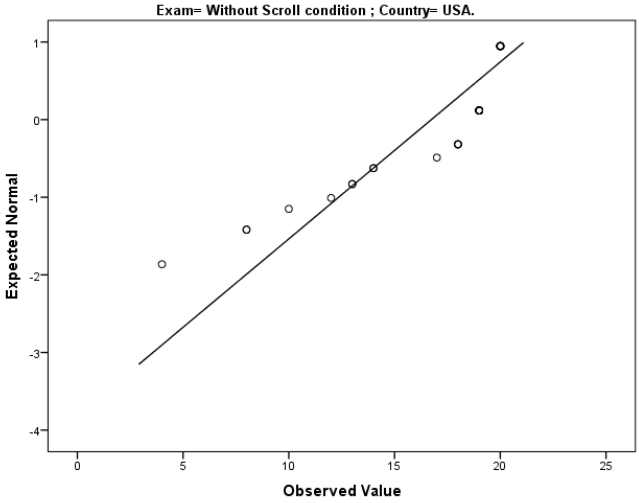


Perceived Ease of Use

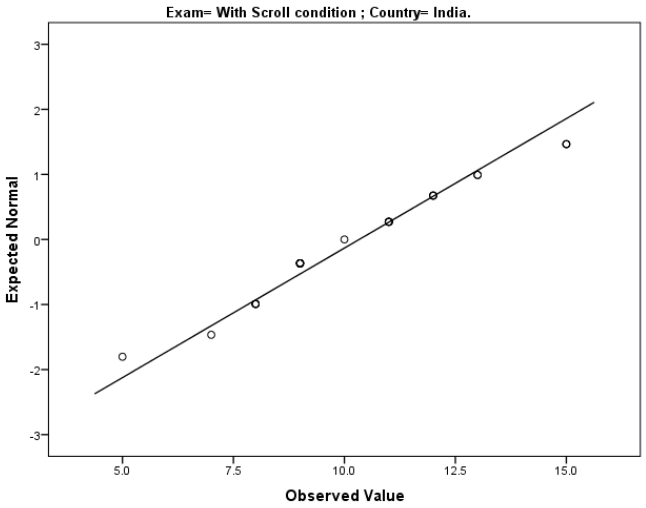


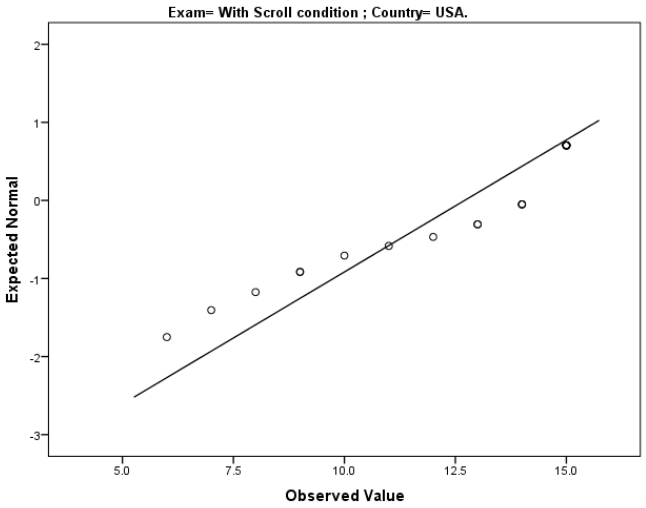


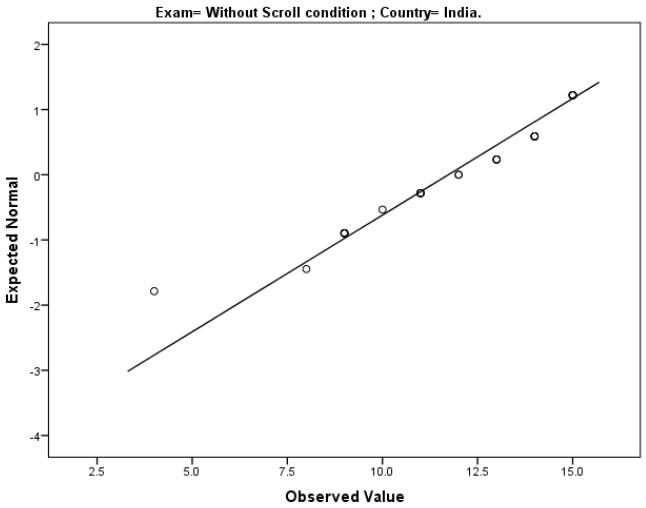


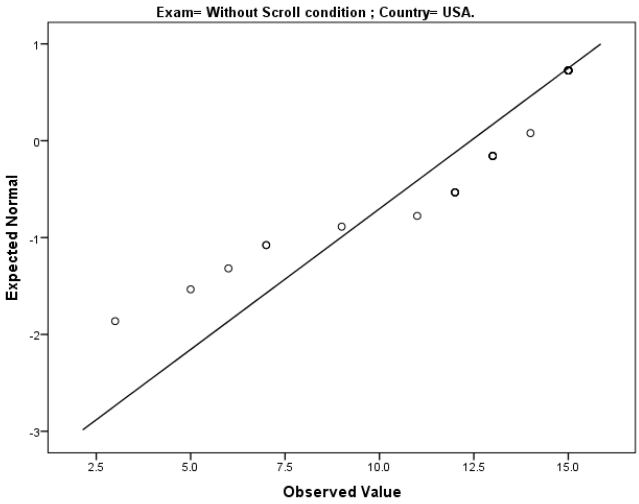


Perceived Usefulness

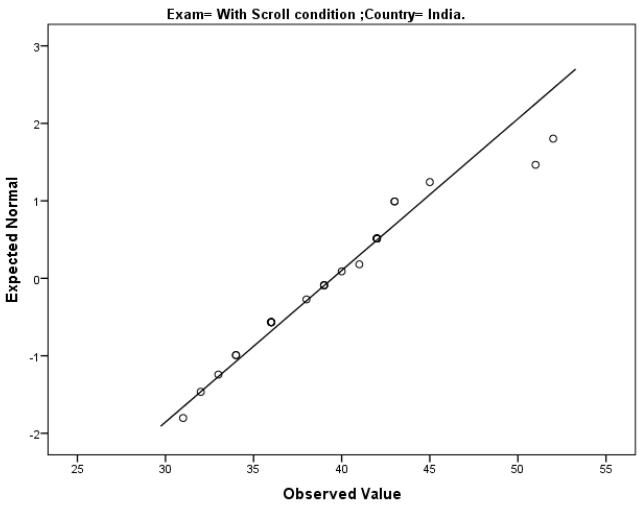


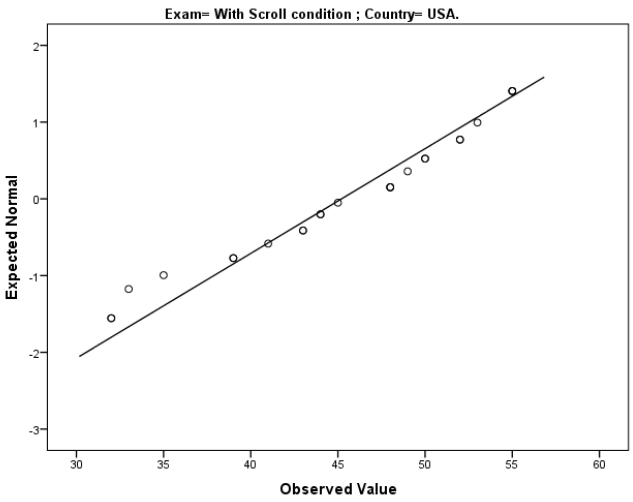


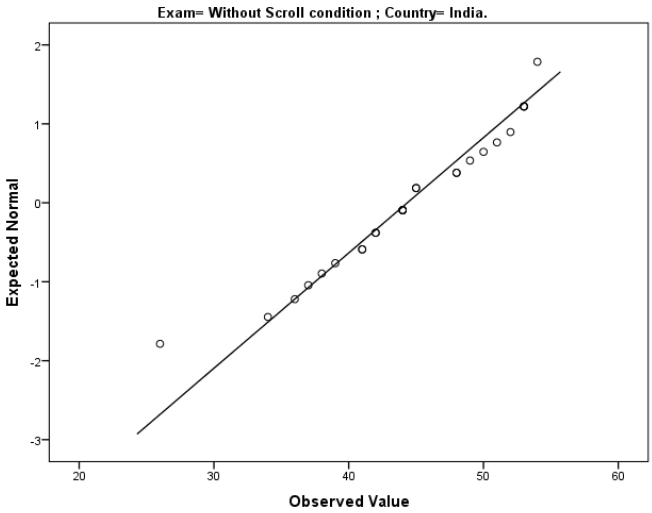


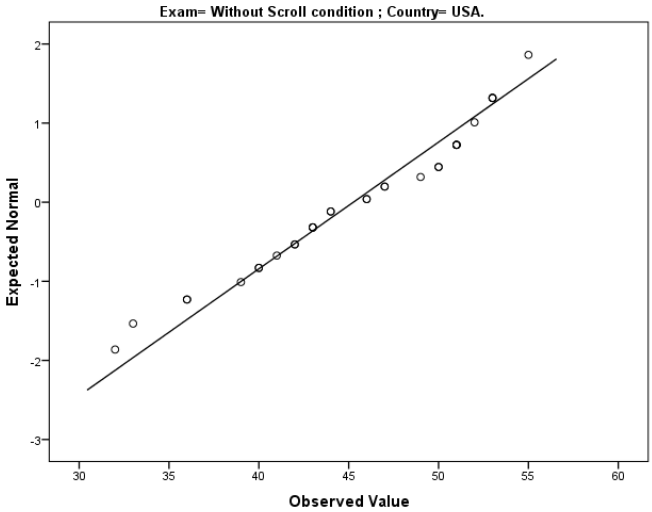


Perceived System Quality

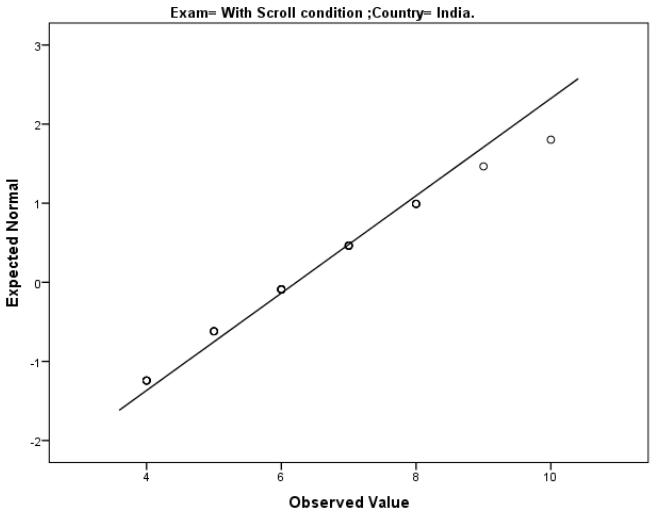


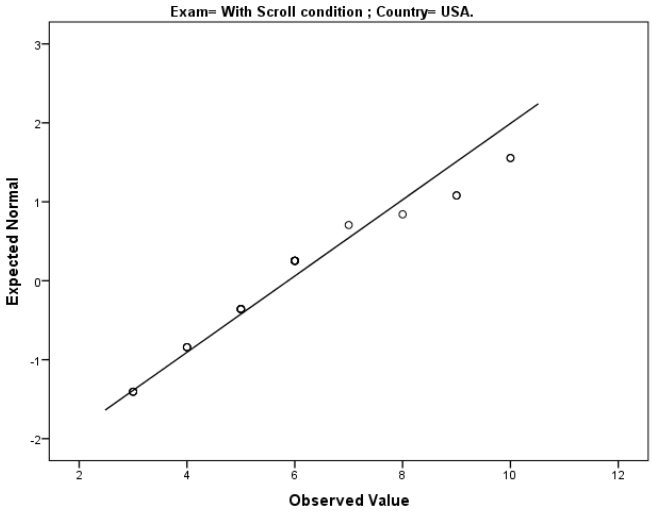


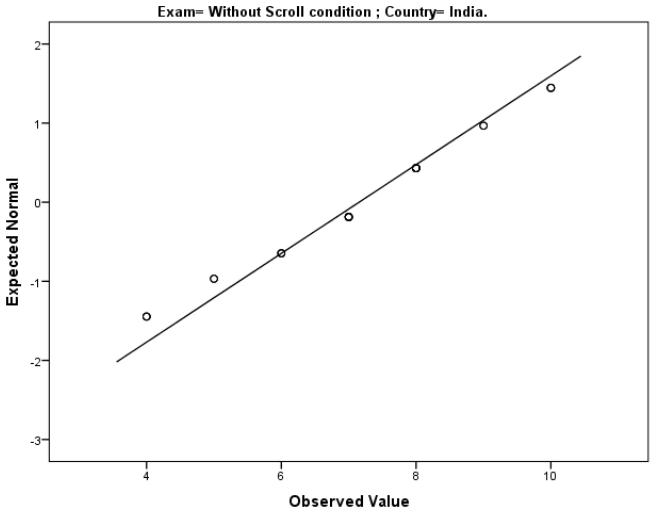


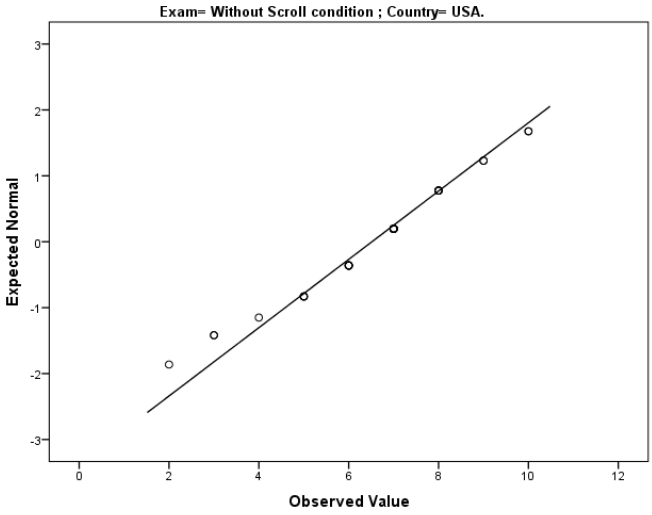


Relative advantage

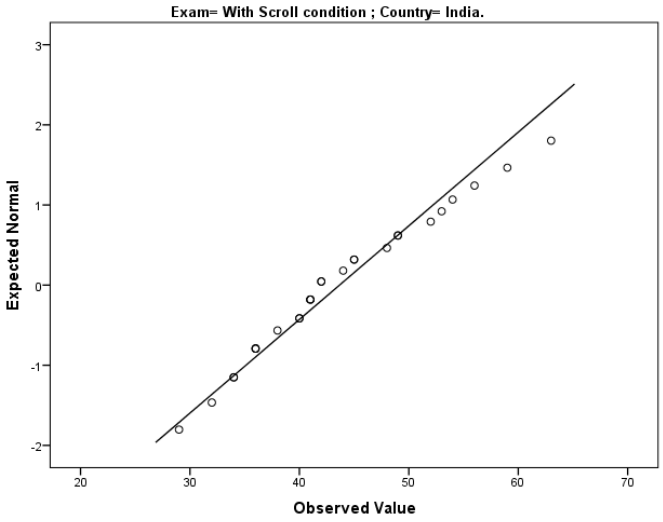


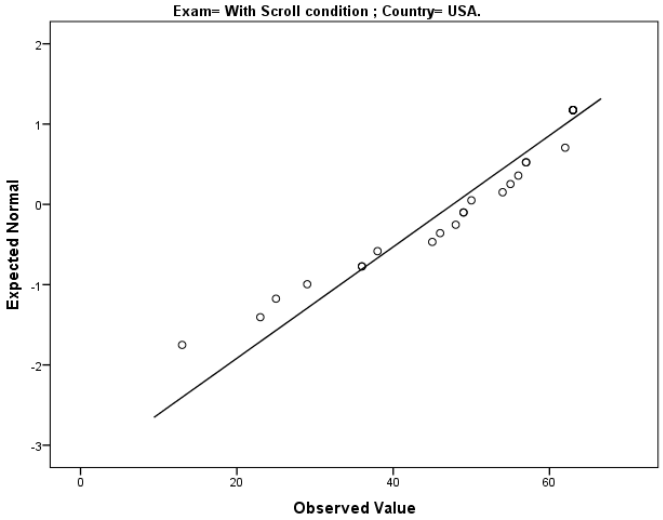


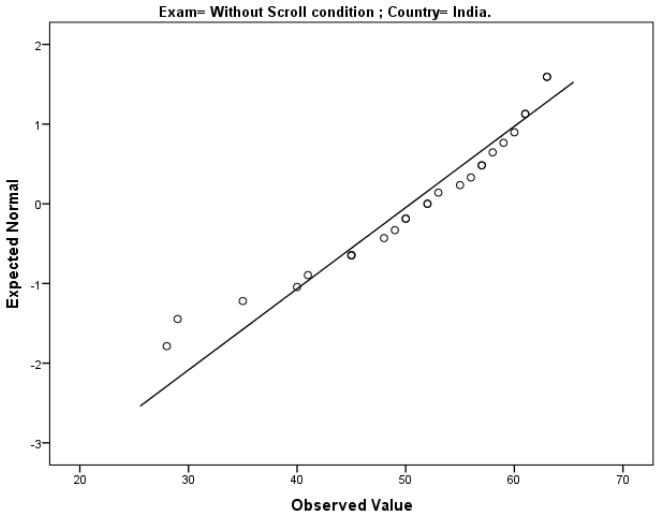


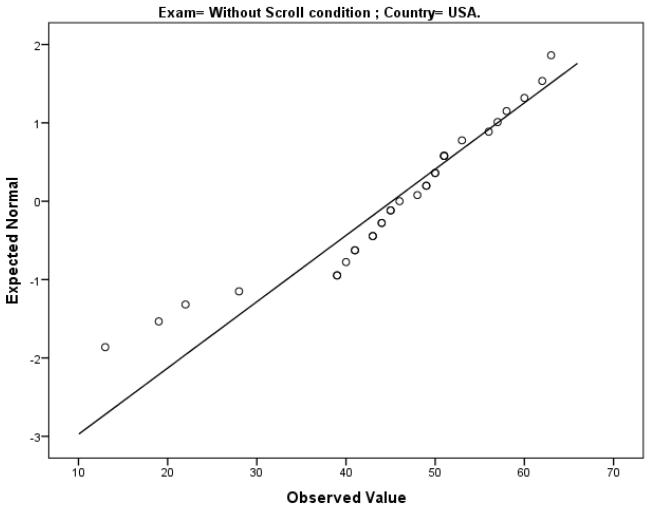


Perceived e-learner satisfaction

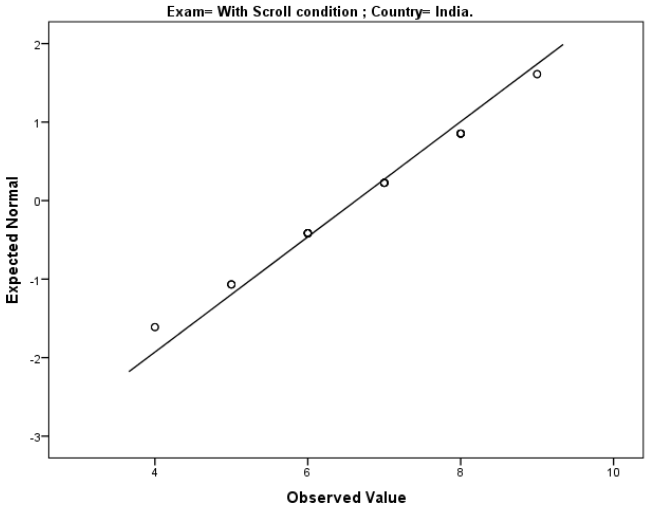


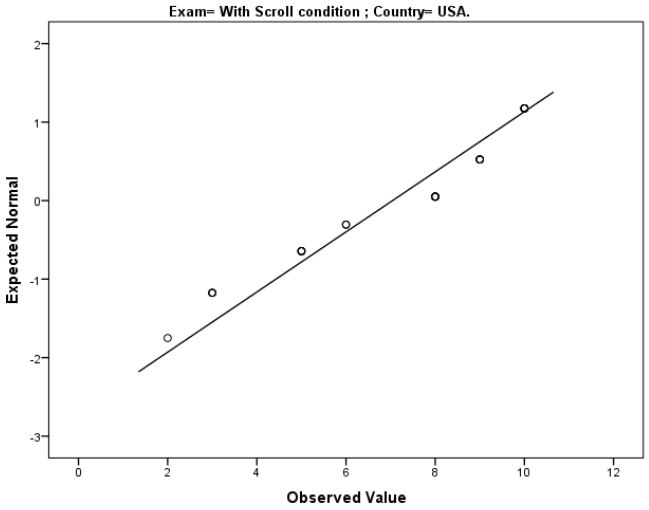


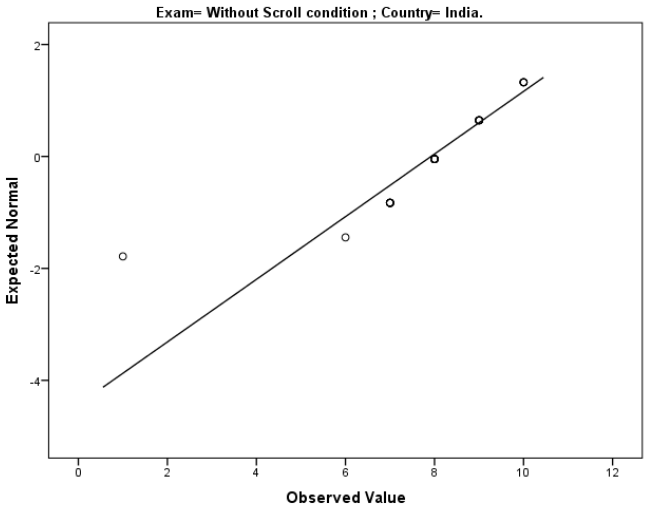


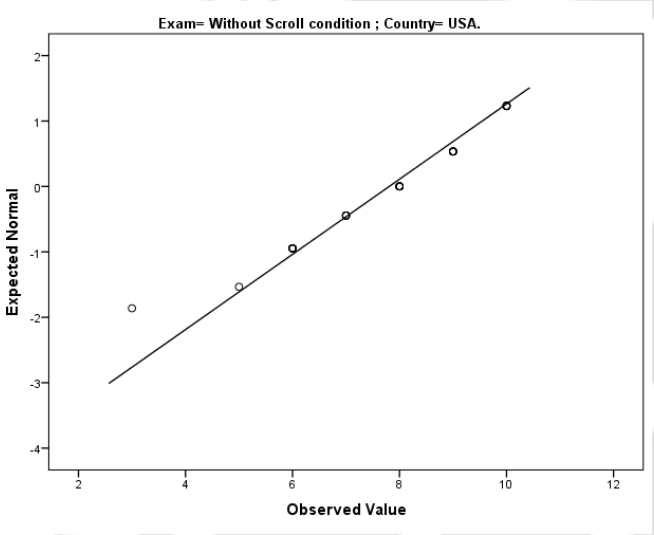


Willingness to continue the usage

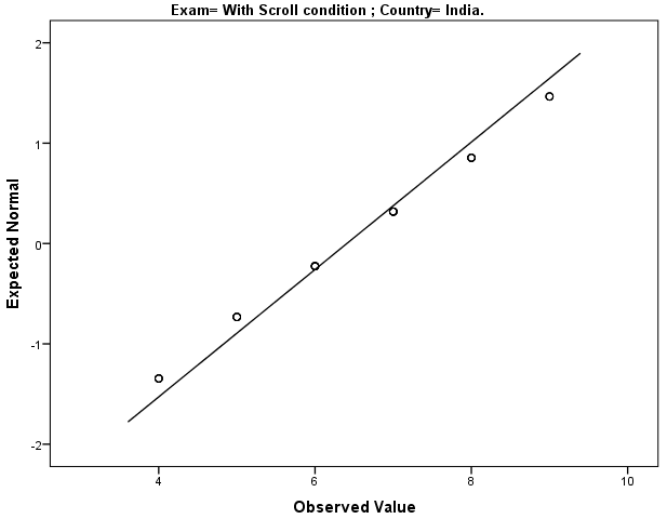


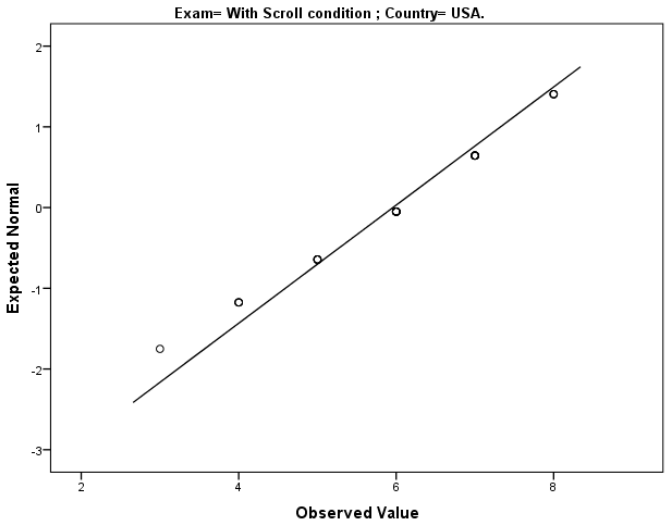


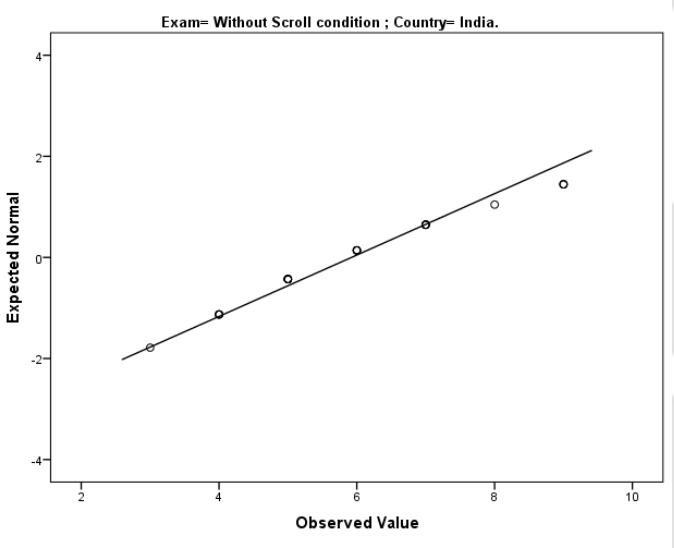


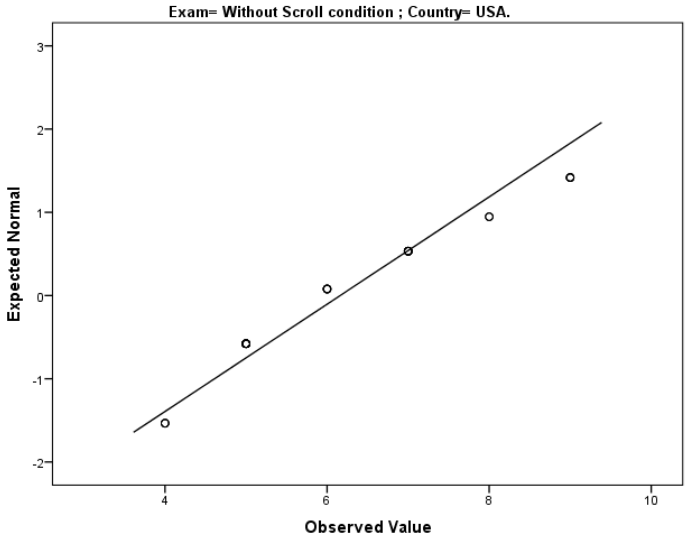


Total Marks obtained

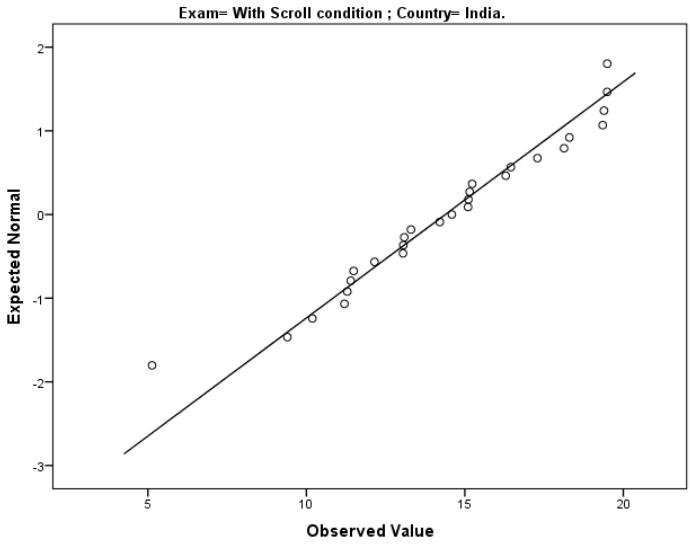


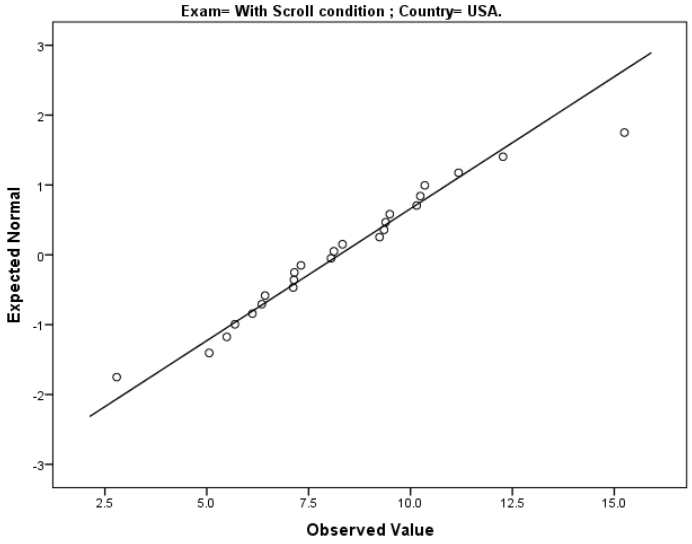


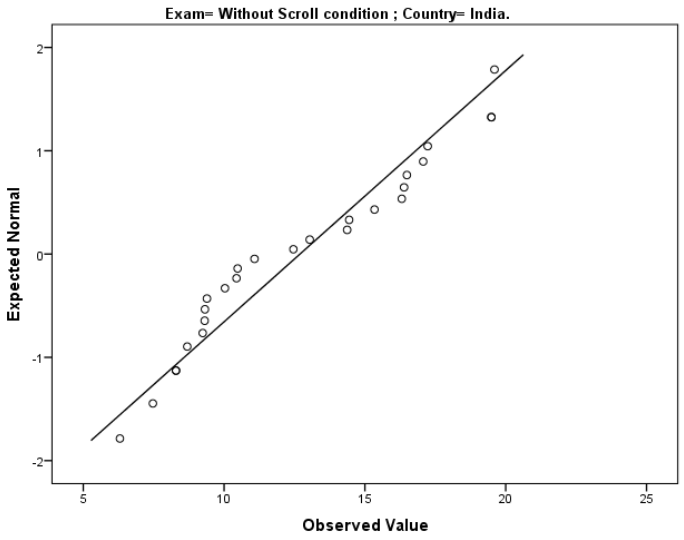




Total time taken







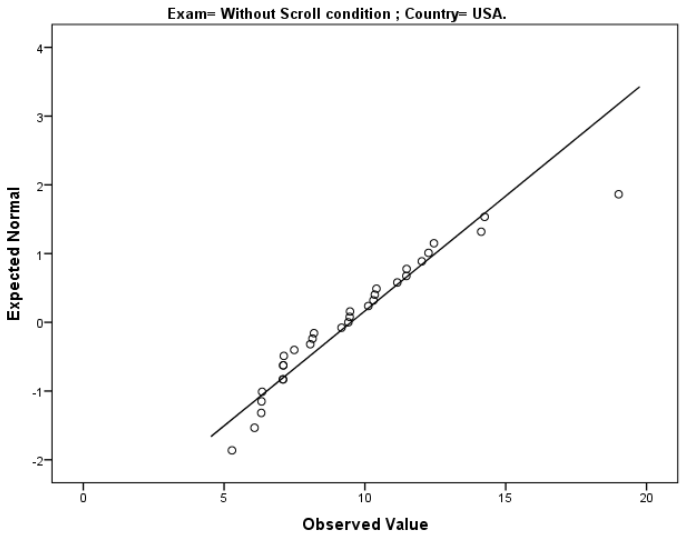


Table 3A.1.

Box's test of equality of covariance matrices.

Box's Test of Equality of Covariance Matrices[a]	
Box's M	235.94
F	1.18
df1	165
df2	23129.06
Sig.	.05

Box's test of equality of covariance matrices highlights that it is not significant as p>0.005. Therefore, the assumption of homogeneity of variance-covariance matrices was not violated.

Table 3A.2

Levene's Test of Equality of Error Variances[a]

	F	df1	df2	Sig.
NASATLX	.76	3	104	.51
SUS	.53	3	104	.65
PEOU	3.32	3	104	.02
PEU	.61	3	104	.60
SQ	1.48	3	104	.22
RA	.23	3	104	.86
PES	1.59	3	104	.19
WCU	4.48	3	104	.00
TM	.38	3	104	.76
TK	3.92	3	104	.01

Table 3A.3

Univariate tests for the main effects of Exam and Country.

Source	Dependent Variable	Type III Sum of	df	Mean Sq.	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[k]
Exam	**NASATLX**	894.56	1	894.56	4.20	.04	.03	4.20	.52
	SUS	.83[a]	1	.83	.00	.94	.00	.00	.05
	PEOU	32.15[b]	1	32.15	2.61	.10	.02	2.61	.36
	PEU	3.83[c]	1	3.83	.43	.51	.00	.43	.10
	SO	103.88[d]	1	103.88	2.54	.11	.02	2.54	.35
	RA	12.82[e]	1	12.82	3.69	.05	.03	3.69	.47
	PES	126.03[f]	1	126.03	.98	.32	.00	.98	.16
	WCU	30.19[g]	1	30.19	8.38	.00	.07	8.38	.81
	TM	.27[h]	1	.27	.11	.73	.00	.11	.06
	TK	.44[i]	1	.44	.03	.84	.00	.03	.05
Countrv	**NASATLX**	376.56[j]	1	376.56	1.76	.18	.01	1.76	.26
	SUS	183.97	1	183.97	.98	.32	.00	.98	.16
	PEOU	20.00	1	20.00	1.62	.20	.01	1.62	.24
	PEU	63.15	1	63.15	7.20	.00	.06	7.20	.75
	SO	302.06	1	302.06	7.40	.00	.06	7.40	.76
	RA	5.79	1	5.79	1.67	.19	.01	1.67	.24
	PES	15.58	1	15.58	.12	.72	.00	.12	.06
	WCU	.81	1	.81	.22	.63	.00	.22	.07
	TM	.27	1	.27	.11	.73	.00	.11	.06
	TK	574.89	1	574.89	51.04	.00	.32	51.04	1.00

Table 3A.4

Pairwise comparisons of the interaction effect.

Dependent Variable	Exam	Country	Mean	Std. Error	95% Confidence Interval	
					Lower Bound	Upper Bound
NASATLX	With Scroll	India	41.24	2.80	35.67	46.81
		USA	35.83	2.86	30.15	41.50
	Without Scroll	India	26.33	2.86	20.65	32.00
		USA	39.21	2.71	33.84	44.59
SUS	With Scroll	India	75.40	2.62	70.19	80.61
		USA	85.00	2.67	79.69	90.30
	Without Scroll	India	82.21	2.67	76.90	87.51
		USA	77.84	2.53	72.81	82.87
PEOU	With Scroll	India	14.66	.67	13.32	16.00
		USA	17.15	.68	15.79	18.51
	Without Scroll	India	17.38	.68	16.02	18.74
		USA	16.62	.65	15.33	17.91
PEU	With Scroll	India	10.33	.57	9.20	11.46
		USA	12.88	.58	11.73	14.03
	Without Scroll	India	11.73	.58	10.58	12.88
		USA	12.24	.55	11.15	13.33
SQ	With Scroll	India	39.48	1.22	37.04	41.91
		USA	45.73	1.25	43.24	48.21
	Without Scroll	India	44.34	1.25	41.86	46.82
		USA	44.79	1.18	42.44	47.14
RA	With Scroll	India	6.22	.35	5.51	6.93
		USA	6.00	.36	5.27	6.72
	Without Scroll	India	7.15	.36	6.42	7.87
		USA	6.44	.34	5.76	7.13

Table 3A.5

Independent sample t-test across the interface features

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
With Scroll	NASATLX	.04	.84	1.49	51	.14	5.41	3.62	-1.86	12.68
	SUS	.08	.78	-2.86	51	.01	-9.59	3.35	-16.33	-2.86
	PEOU	1.46	.23	-2.87	51	.01	-2.49	.87	-4.23	-.75
	SQ	4.61	.04	-3.62	51	.00	-6.25	1.73	-9.72	-2.78
	PES	4.57	.04	-1.23	51	.22	-3.87	3.15	-10.20	2.45
	TK	3.01	.09	7.26	51	.00	6.18	.85	4.47	7.89
Without Scroll	NASATLX	.39	.53	-2.99	54	.00	-12.57	4.21	-21.00	-4.14
	SUS	.09	.77	1.15	54	.25	4.56	3.95	-3.36	12.49
	PEOU	6.26	.02	.85	54	.40	.86	1.01	-1.17	2.90
	SQ	.09	.77	-.09	54	.93	-.16	1.74	-3.65	3.32
	PES	.50	.48	1.82	54	.07	5.38	2.95	-.54	11.29
	TK	6.93	.01	3.34	54	.00	3.22	.96	1.29	5.16

Table 3A.5

Independent sample t-test across the cultural backgrounds

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
India	NASATLX	.44	.51	3.74	52	.00	14.59	3.90	6.76	22.42
	SUS	.32	.58	-1.96	52	.06	-7.00	3.58	-14.18	0.18
	PEOU	.02	.89	-3.69	52	.00	-2.81	.76	-4.35	-1.28
	SQ	2.02	.16	-3.12	52	.00	-5.15	1.65	-8.46	-1.84
	PES	.15	.70	-2.74	52	.01	-6.78	2.48	-11.75	-1.81
	TK	2.49	.12	1.46	52	.15	1.52	1.05	-.57	3.62
USA	NASATLX	1.52	.22	-.85	53	.40	-3.39	3.98	-11.38	4.60
	SUS	1.07	.30	1.89	53	.06	7.16	3.78	-.44	14.75
	PEOU	2.03	.16	.49	53	.63	.53	1.10	-1.67	2.74
	SQ	.99	.32	.52	53	.61	.94	1.81	-2.70	4.58
	PES	.84	.36	.70	53	.49	2.47	3.51	-4.58	9.52
	TK	.82	.37	-1.88	53	.07	-1.43	.76	-2.96	0.10

Table 3A.6

Correlation of the dependent variables

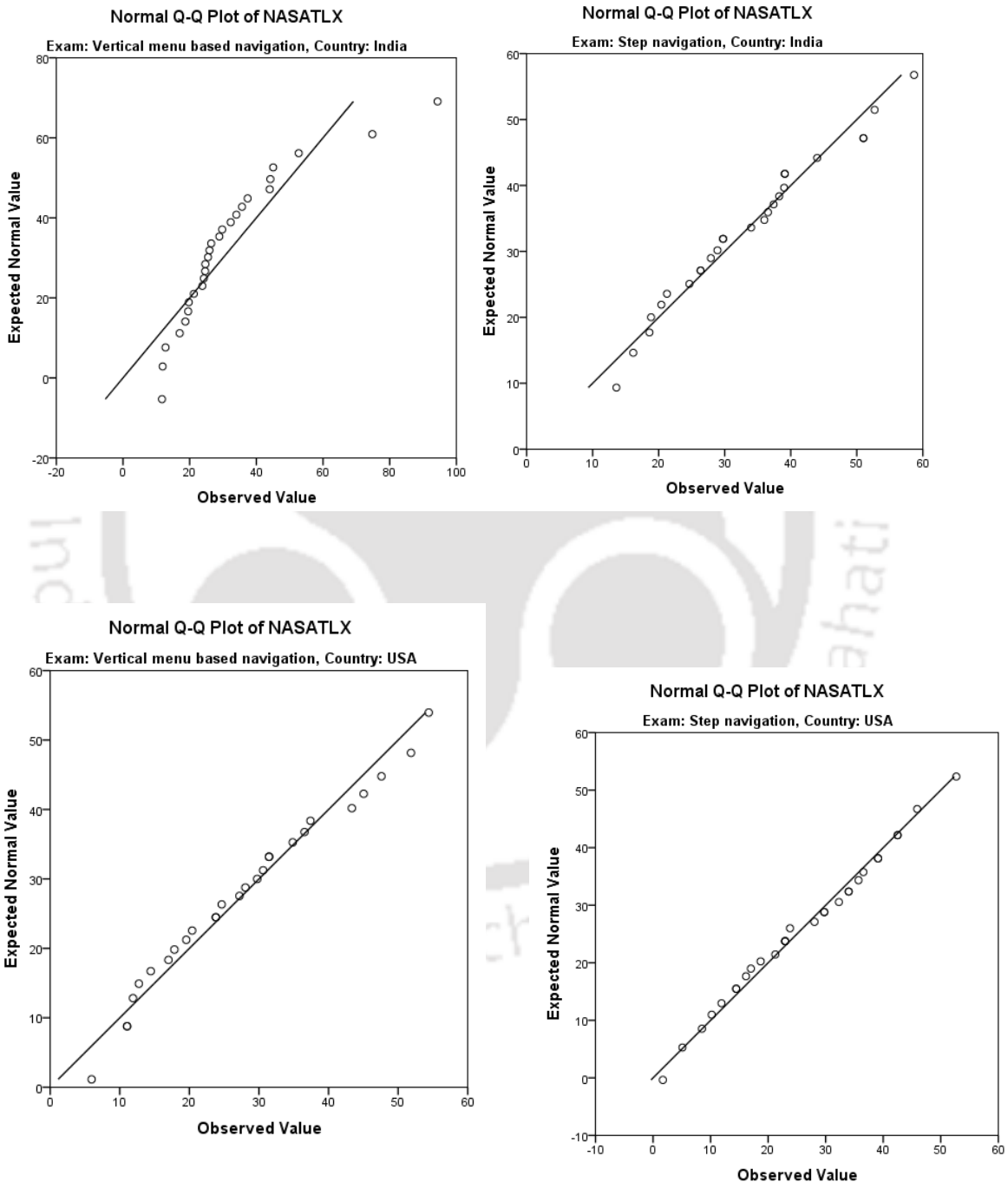
Construct	1	2	3	4	5	6	7	8	9	10
1. NASATLX	1									
2. SUS	-.29**	1								
3. PEOU	-.38**	.35**	1							
4. PEU	-.28**	.40**	.80**	1						
5. SQ	-.35**	.41**	.60**	.61**	1					
6. RA	-.18	.24*	.36**	.42**	.46**	1				
7. PES	-.34**	.50**	.63**	.68**	.54**	.53**	1			
8. WCU	-.11	.22*	.37**	.49**	.44**	.61**	.64**	1		
9. TM	.01	.07	.15	.12	-.14	-.10	.11	.07	1	
10. TK	-.04	.05	-.10	-.13	-.21	-.00	-.00	-.01	.16	1

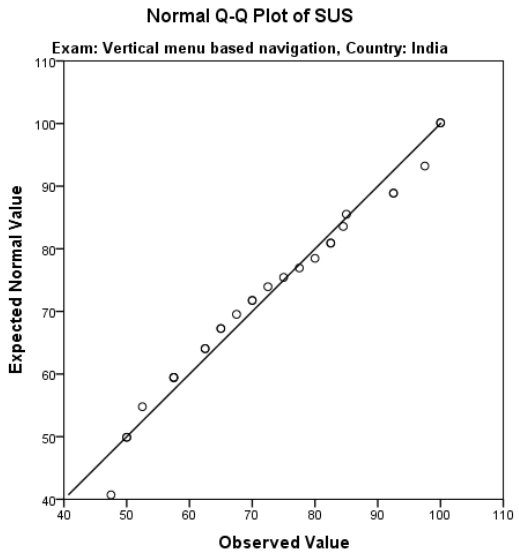
**p<0.01 (2-tailed), *p<0.05 (2-tailed).

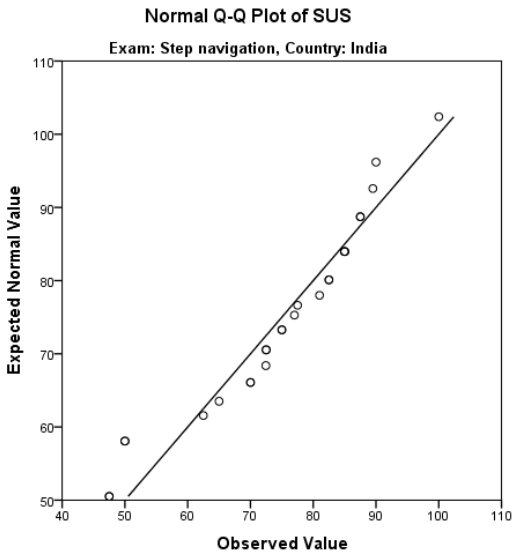
Appendix – 3B

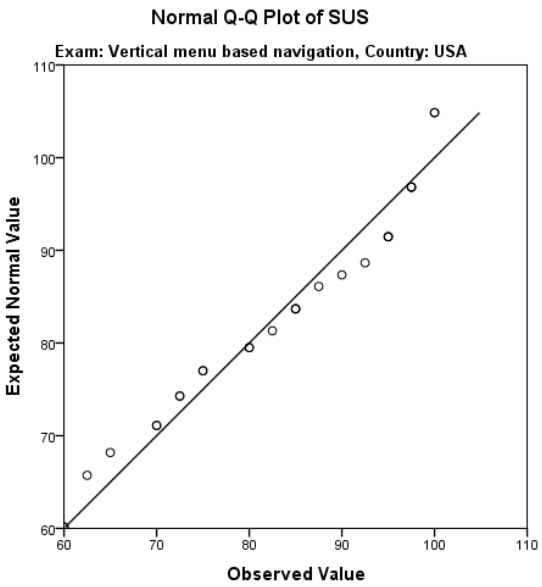
Experiment 2: Vertical menu vs. step based navigation

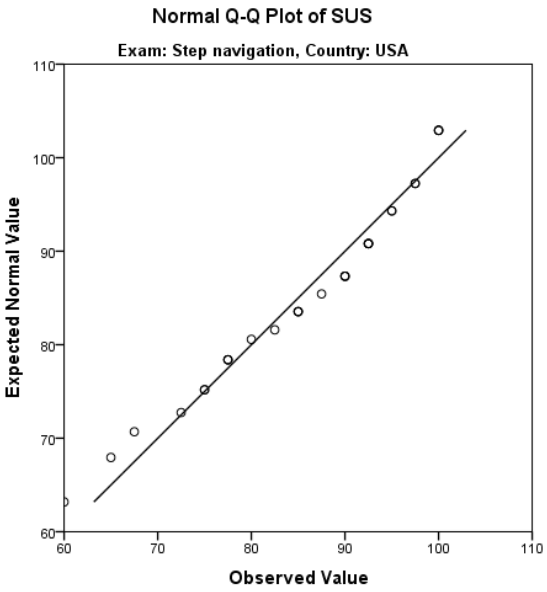
Figure 3B.1 Normal Q-Q plots

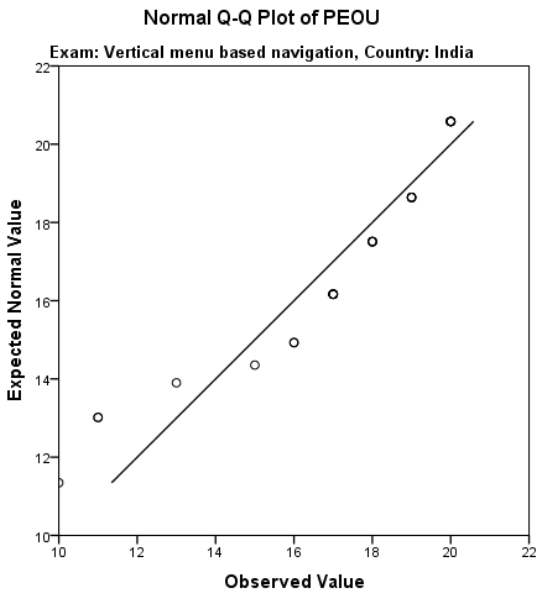


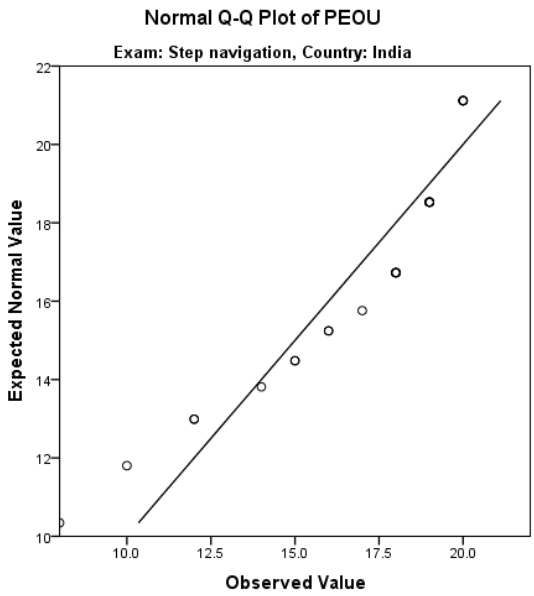


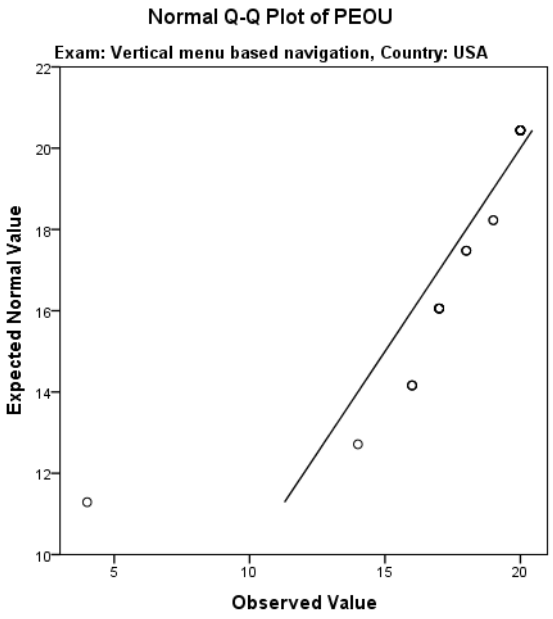


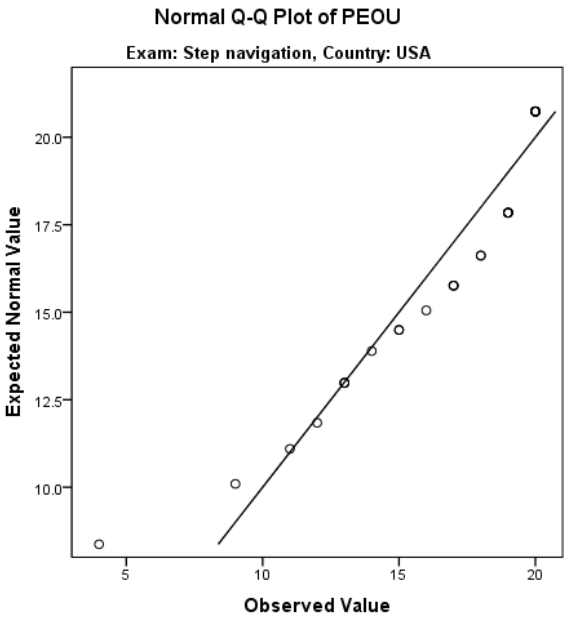


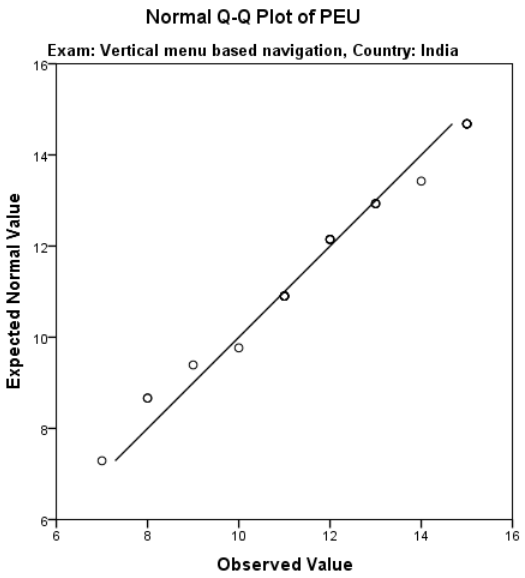


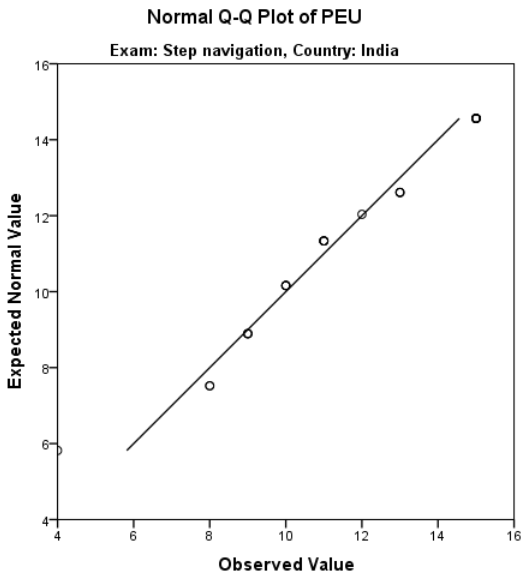


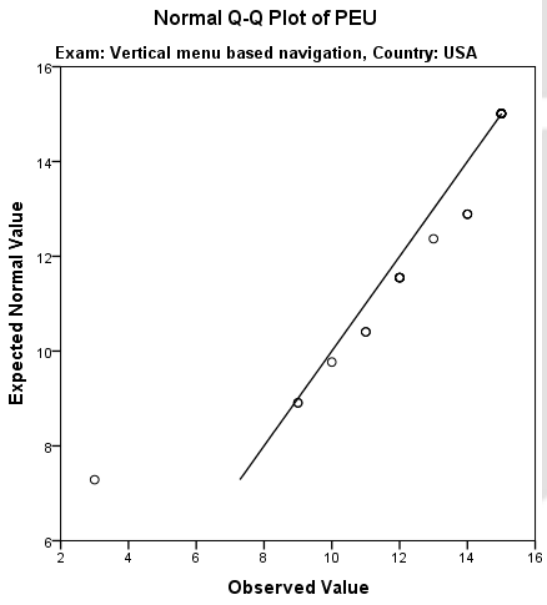


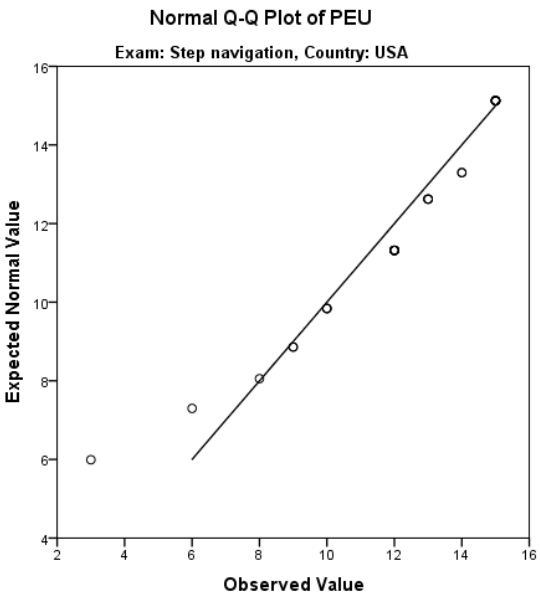


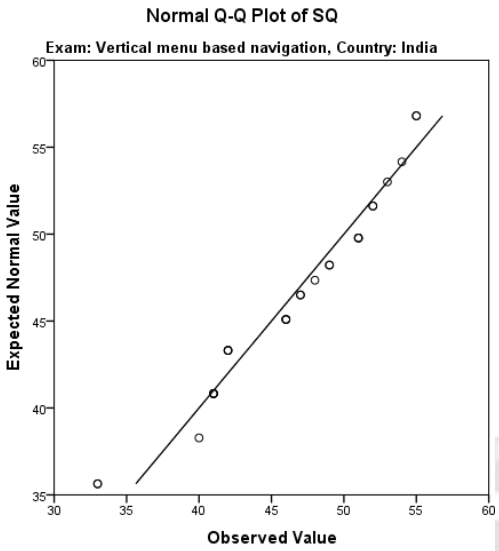


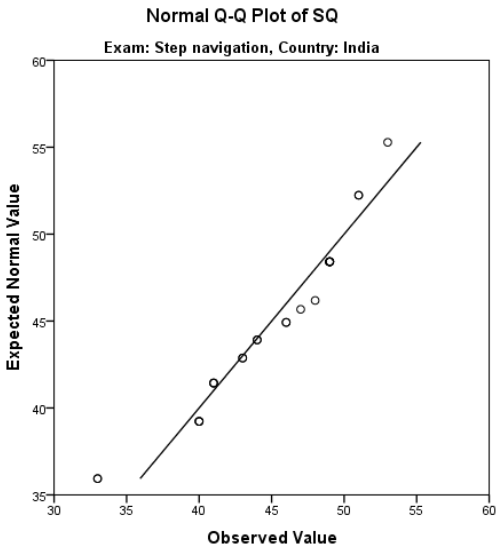


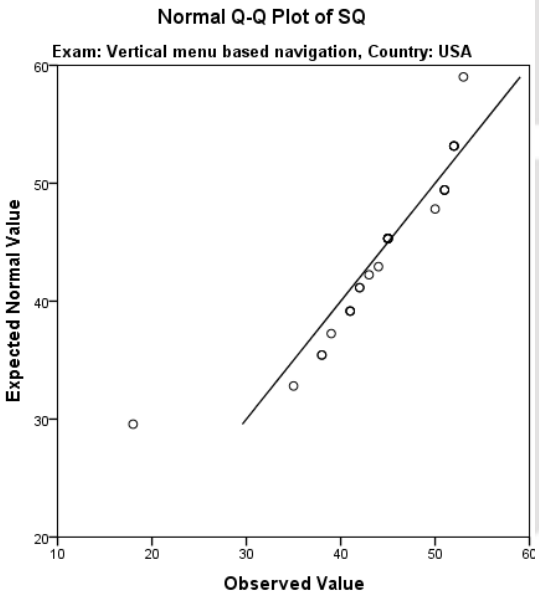


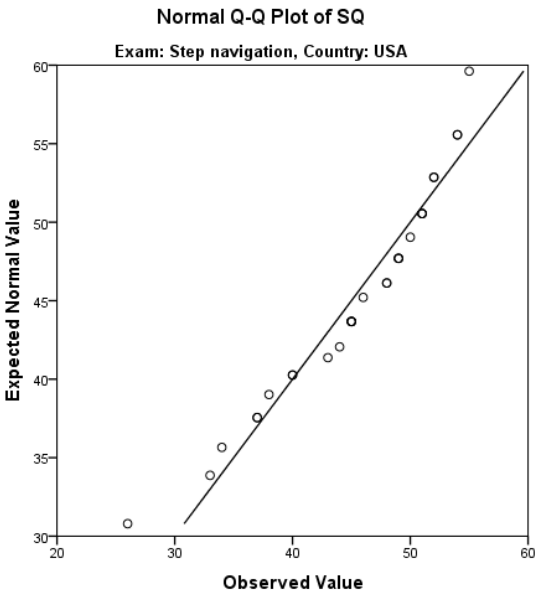


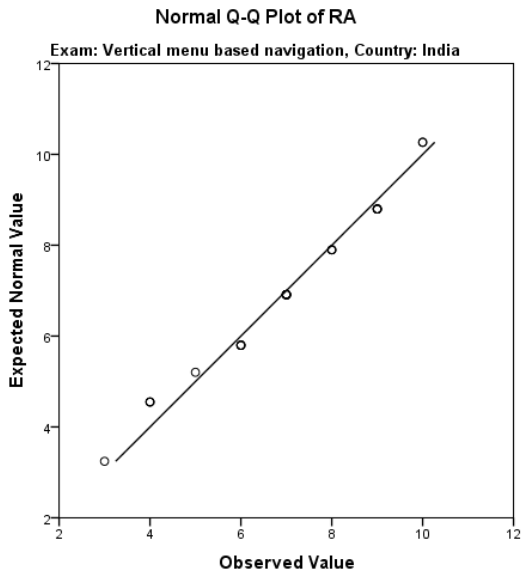


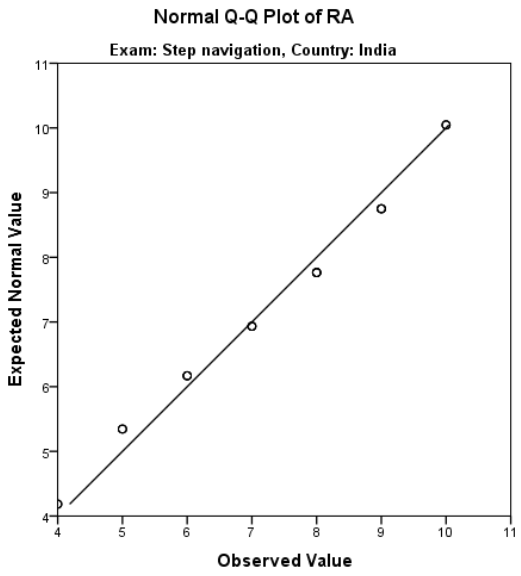


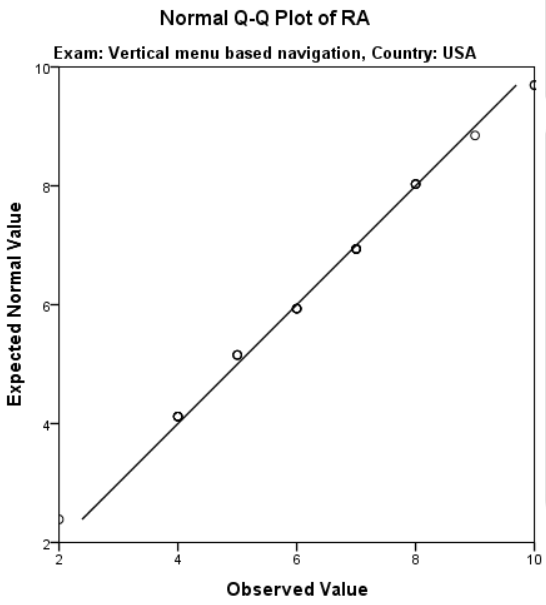


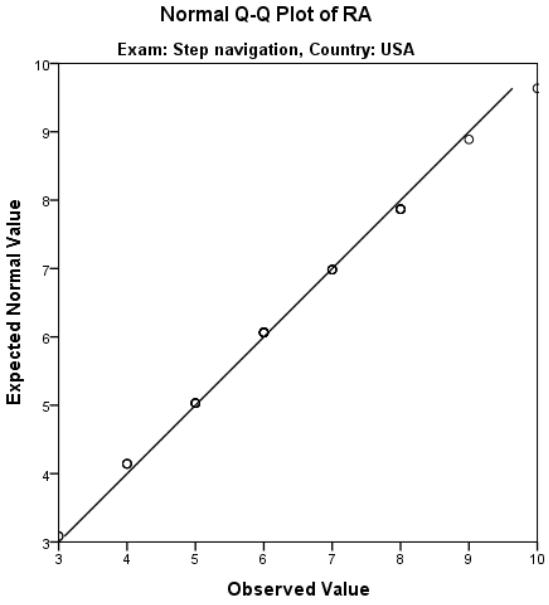


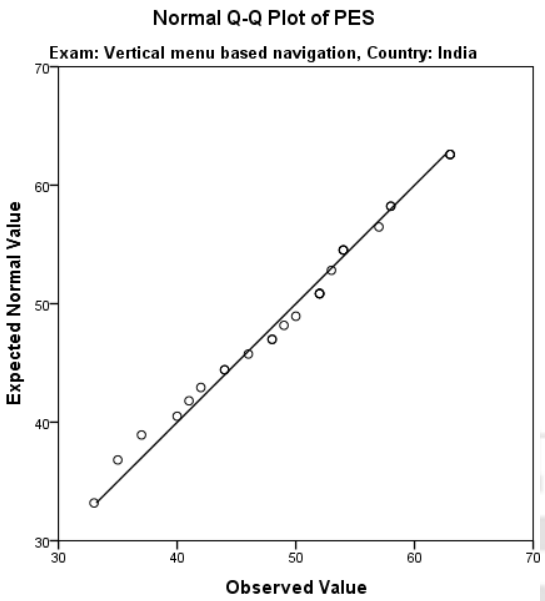


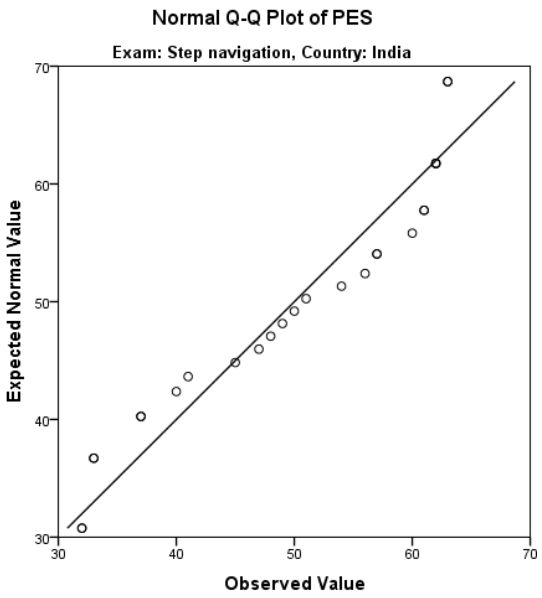


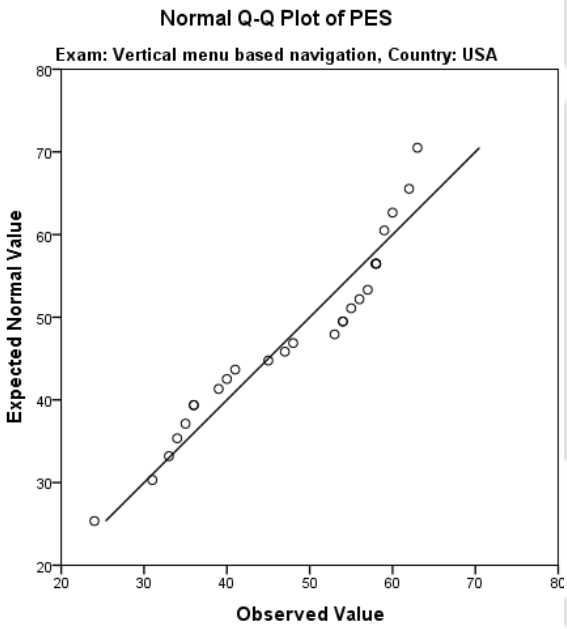


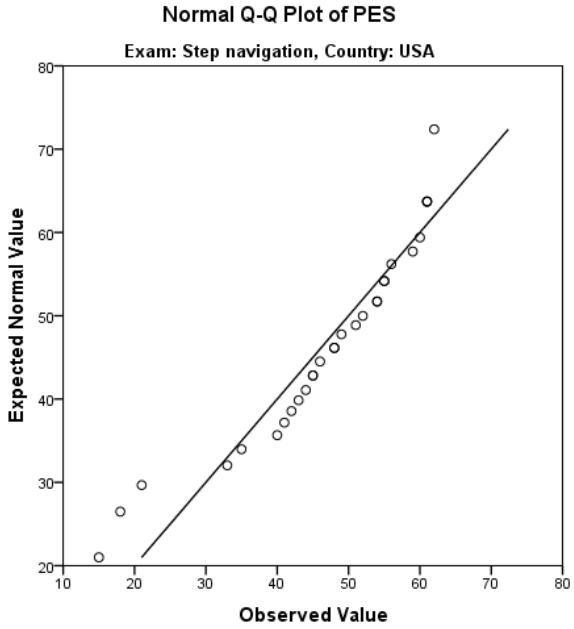


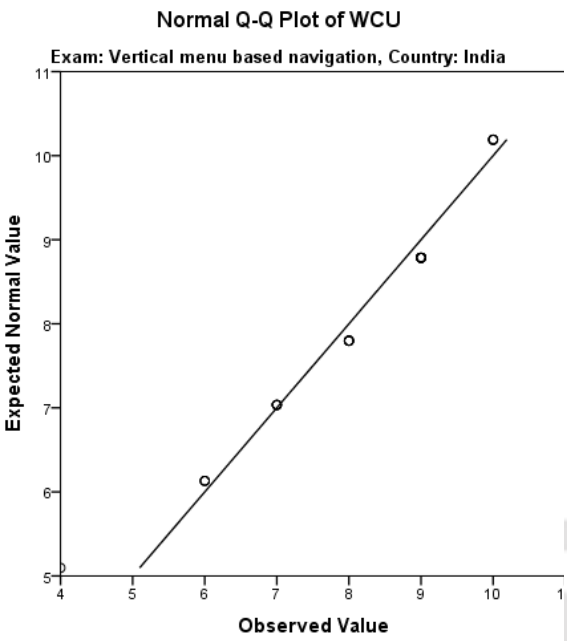


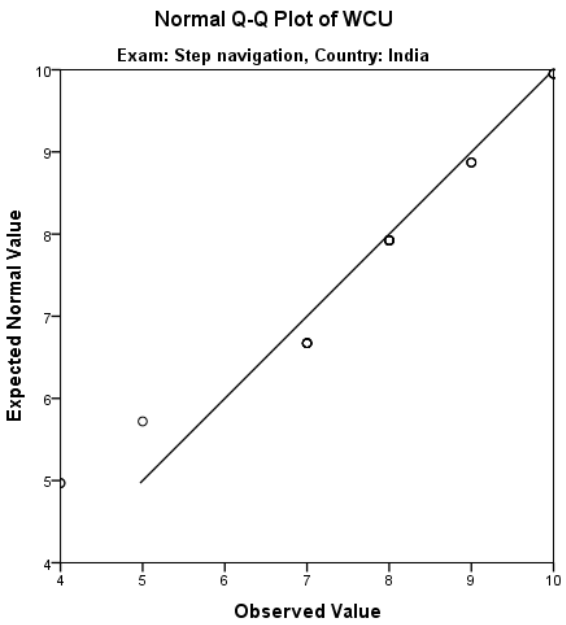


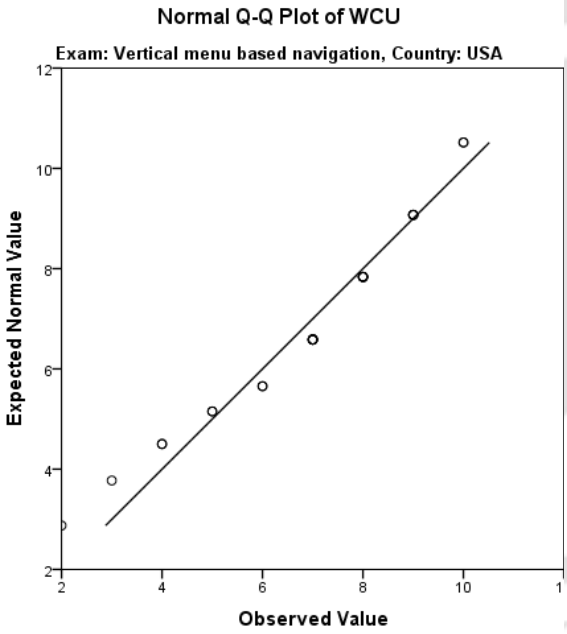


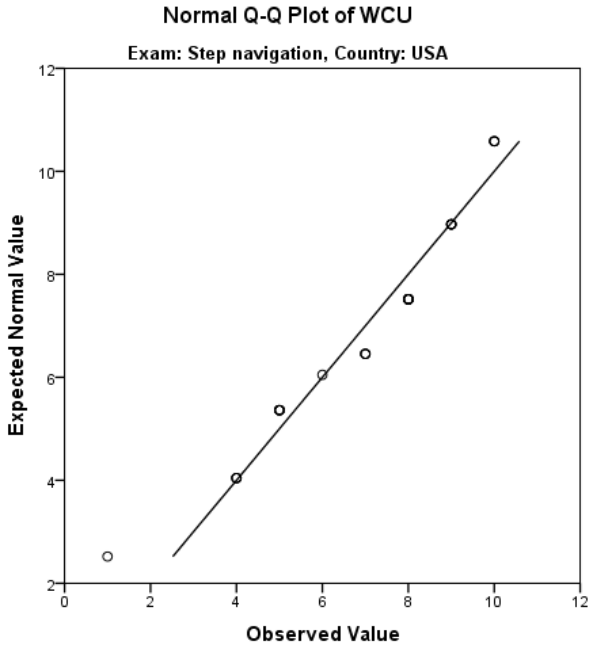


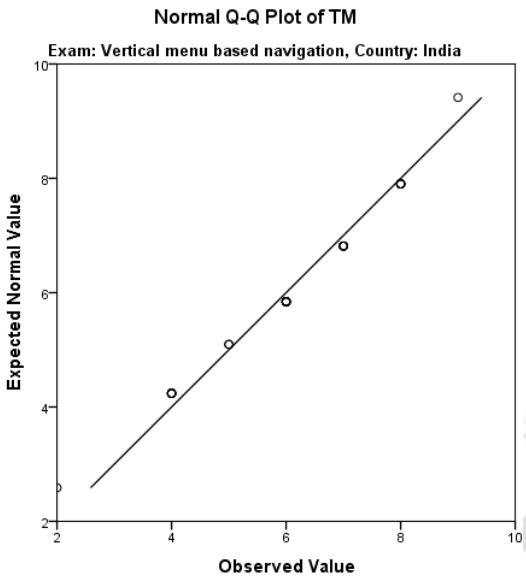


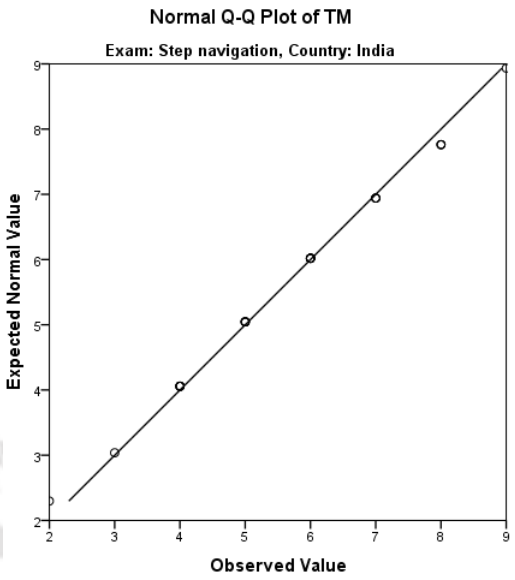


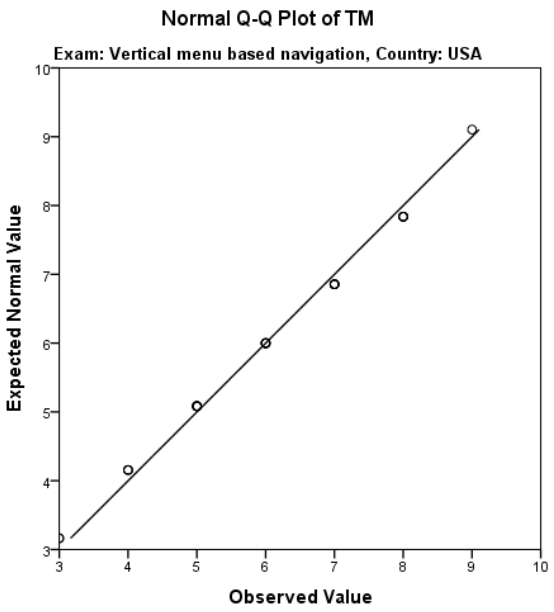


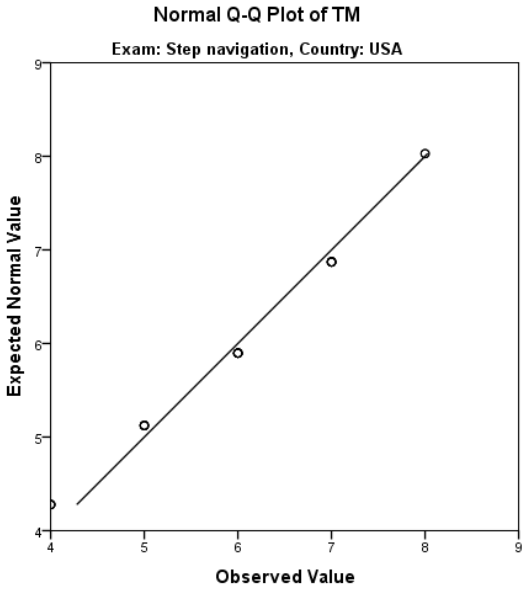


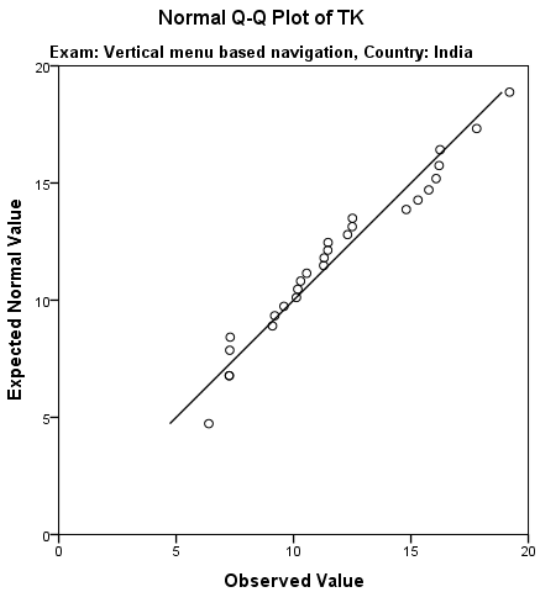


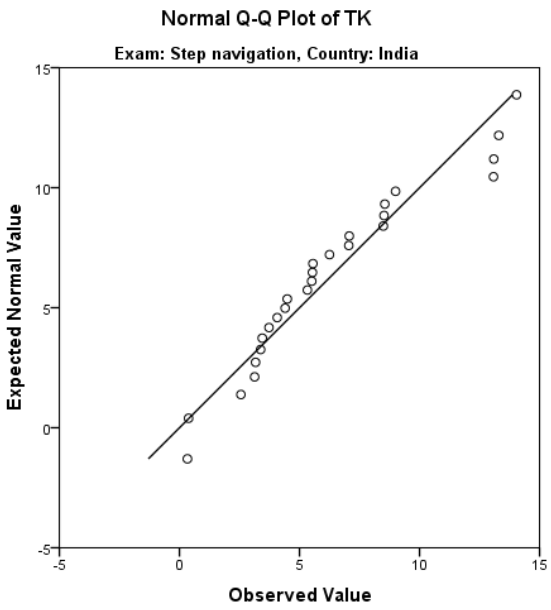


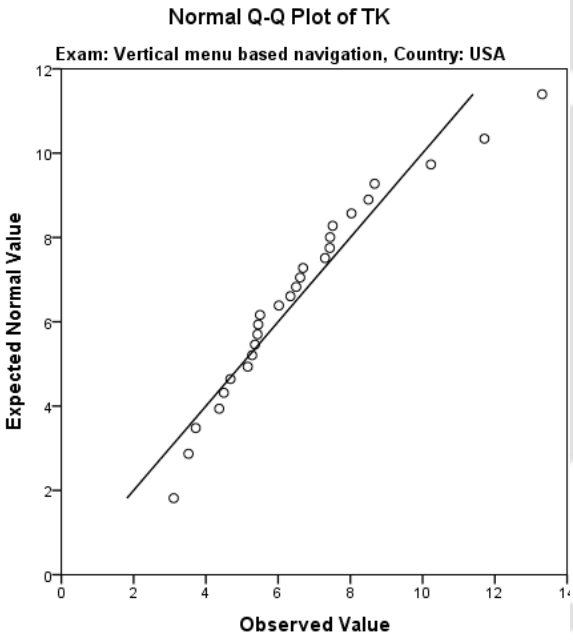












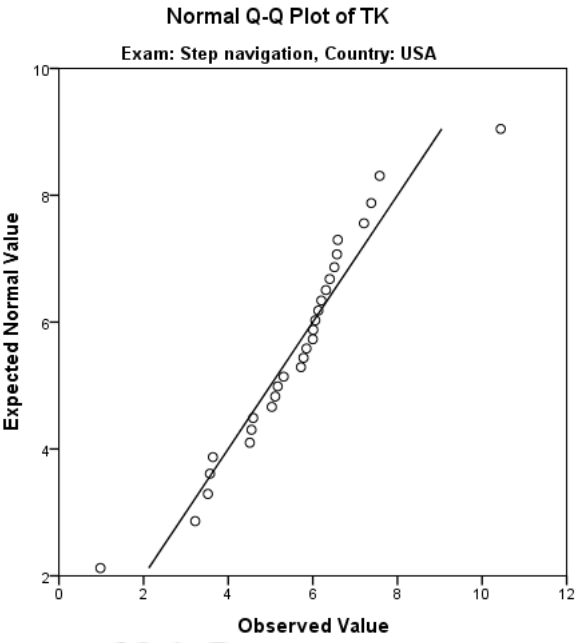


Table 3B.1

Box's test of equality of covariance matrices.

Box's M	267.85
F	1.34
df1	165.00
df2	23643.07
Sig.	.00

Box's test of equality of covariance matrices highlights that it is significant at $p<0.005$. Therefore, the assumption of homogeneity of variance-covariance of matrices has been violated. It suggests that the variance-covariance matrices among the dependent variables across the groups are not the same.

Table 3B.2

Levene's Test of Equality of Error Variances[a] .

	F	*df1*	*df2*	Sig.
NASATLX	.63	3.00	105.00	.60
SUS	1.56	3.00	105.00	.20
PEOU	1.59	3.00	105.00	.20
PEU	.47	3.00	105.00	.71
SQ	.46	3.00	105.00	.71
RA	.32	3.00	105.00	.81
PES	1.74	3.00	105.00	.16
WCU	1.76	3.00	105.00	.16
TM	1.08	3.00	105.00	.36
TK	6.51	3.00	105.00	.00

Levene's test of equality of variances highlights that it is not significant for the dependent measures except TK at $p<0.05$ Therefore, the assumption of homogeneity of variances assumption has been violated for TK.

Table 3B.3

Univariate statistics of the dependent measures on the interacting Independent measures

Source	Dependent Variable	Type III Sum of Squares	*df*	Mean Square	*F*	Sig.	Partial Eta Squared
Exam * Country	**NASATLX**	49.81	1	49.81	.24	.63	.00
	SUS	4.34	1	4.34	.02	.88	.00
	PEOU	8.28	1	8.28	.72	.40	.00
	PEU	.90	1	.90	.12	.73	.00
	SQ	33.21	1	33.21	.78	.38	.00
	RA	.51	1	.51	.15	.70	.00
	PES	10.86	1	10.86	.09	.76	.00
	WCU	.34	1	.34	.09	.76	.00
	TM	.86	1	.86	.36	.55	.00
	TK	137.45	1	137.45	15.62	.00	.12

Table 3B.4a

Pairwise comparisons of time taken between the Exam conditions and the countries.

			Pairwise Comparisons				
			Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
						Lower Bound	Upper Bound
TK	Vertical menu based navigation	Step navigation	3.27*	.56	.00	2.14	4.39
	Step navigation	Vertical menu based navigation	-3.27*	.56	.00	-4.39	-2.14

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Table 3B.4b

Pairwise comparisons of the dependent measures between the Exam conditions and the countries.

Dependent Variable	(I) Country	(J) Country	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
						Lower Bound	Upper Bound
NASATLX	India	USA	5.69*	2.76	.04	.20	11.18
	USA	India	-5.69*	2.76	.04	-11.18	-.20
SUS	India	USA	-10.41*	2.59	.00	-15.55	-5.27
	USA	India	10.41*	2.59	.00	5.27	15.55
RA	India	USA	.84*	.35	.02	.13	1.55
	USA	India	-.84*	.35	.02	-1.55	-.13
WCU	India	USA	1.00*	.36	.00	.27	1.72
	USA	India	-1.00*	.36	.00	-1.72	-.27
TK	India	USA	2.95*	.56	.00	1.82	4.07
	USA	India	-2.95*	.56	.00	-4.07	-1.82

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Table 3B.5

Correlation of the dependent variables

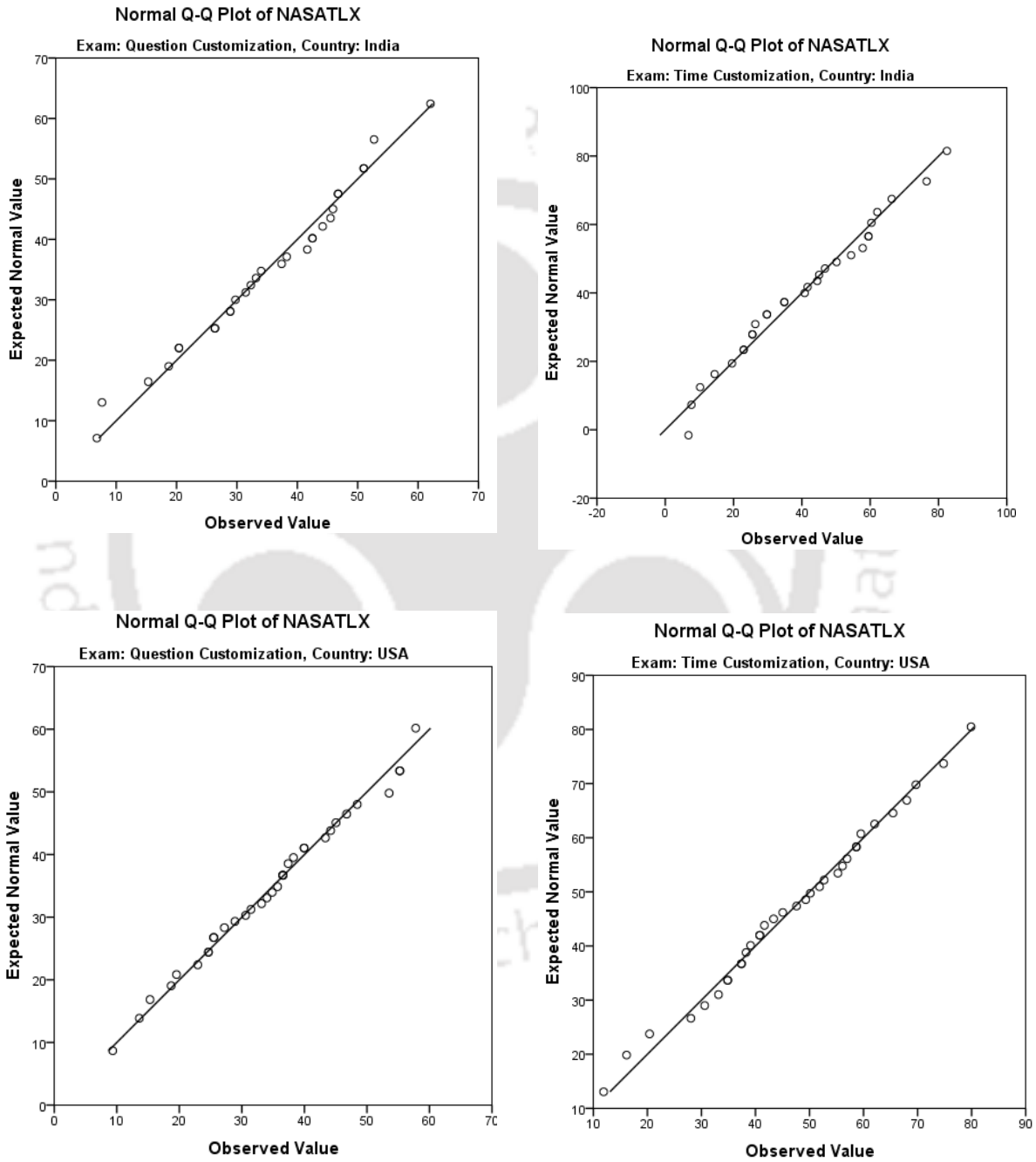
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1. NASATLX	1									
2. SUS	-.190*	1								
3. PEOU	-.218*	.373**	1							
4. PEU	-.237*	.286**	.675**	1						
5. SQ	-.033	.305**	.558**	.511**	1					
6. RA	.111	.169	.423**	.271**	.370**	1				
7. PES	-.133	.444**	.525**	.543**	.457**	.461**	1			
8. WCU	.012	.182	.494**	.454**	.462**	.502**	.599**	1		
9.TM	-.048	.243*	.158	.087	.106	-.040	.163	.237*	1	
10. TK	.129	-.237*	-.118	.001	.082	.044	.038	.062	.025	1

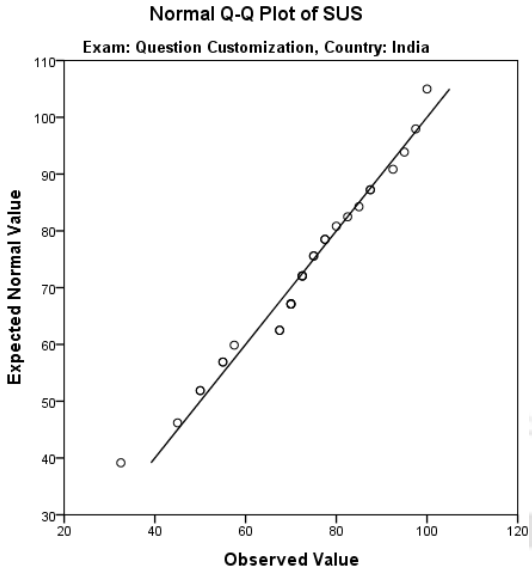
*$p < 0.05$ (2-tailed); **$p < 0.01$ (2-tailed).

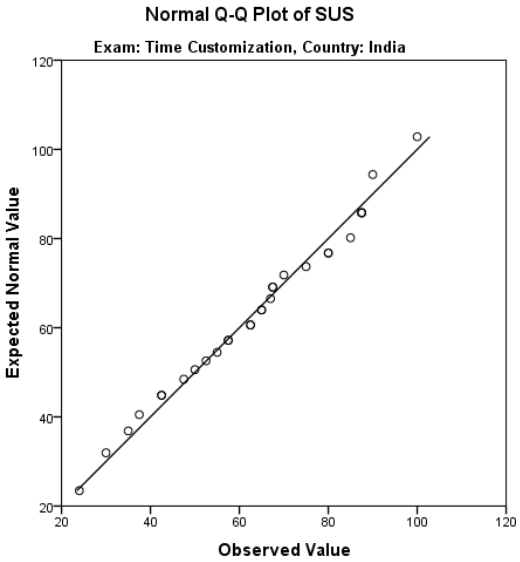
Appendix – 3C

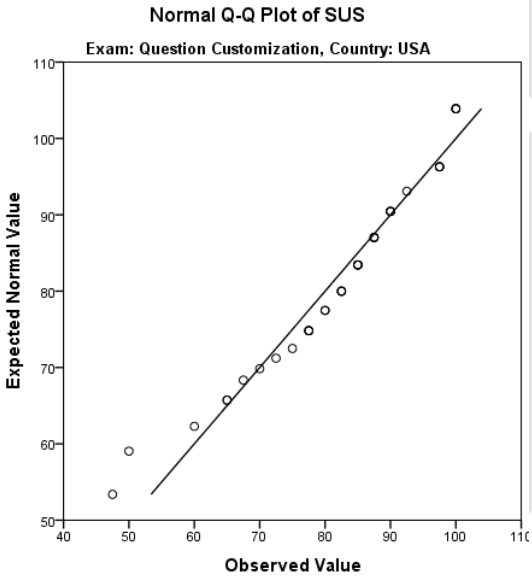
Experiment 3: Time Customization vs. Question Customization

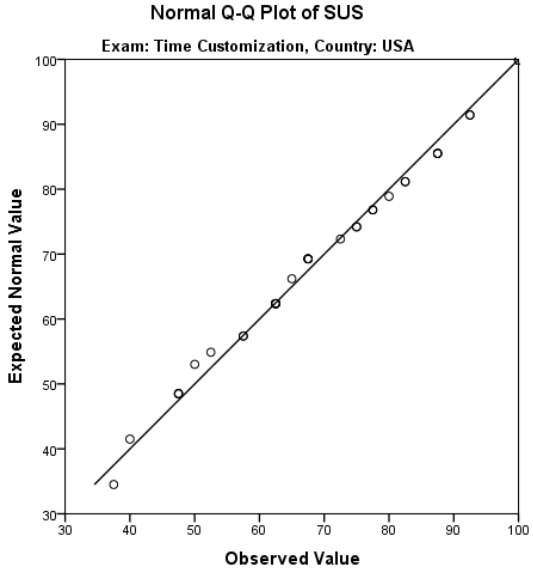
Figure 3C.1 Normal Q-Q plots

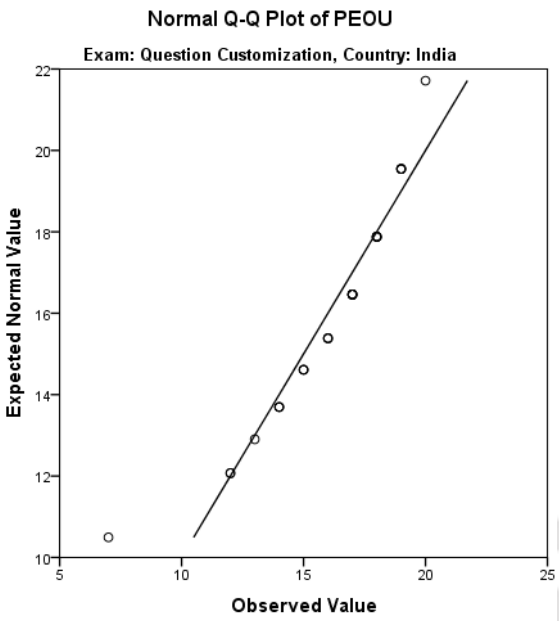


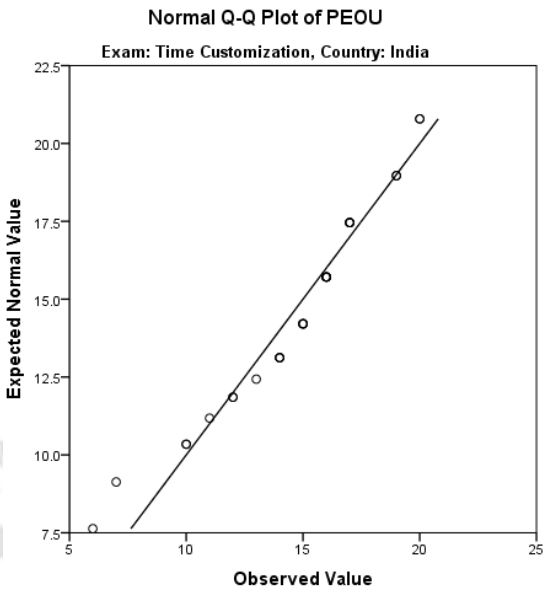


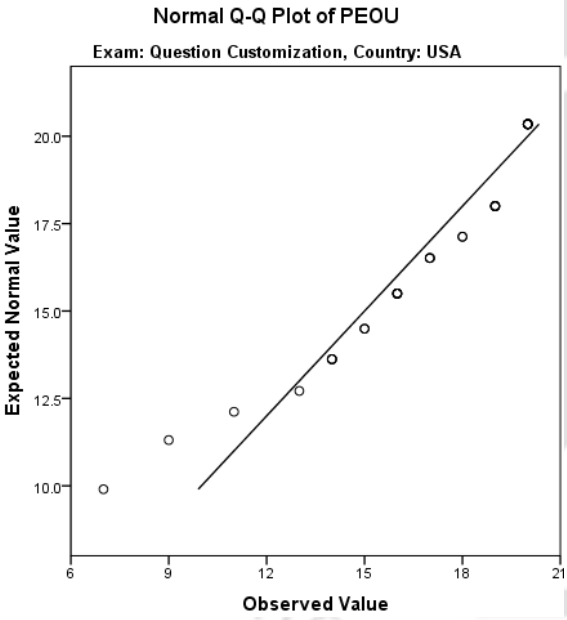


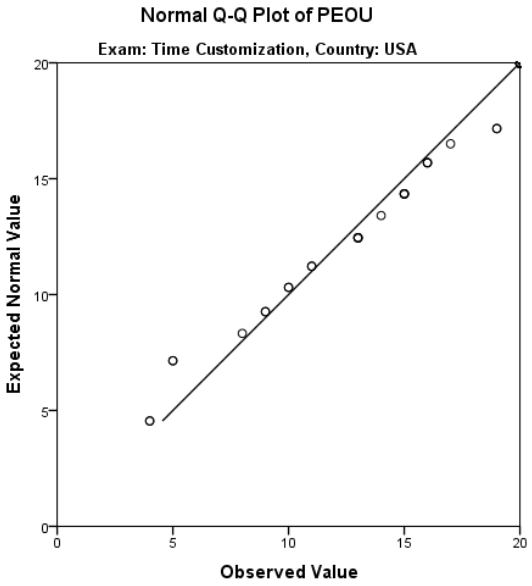


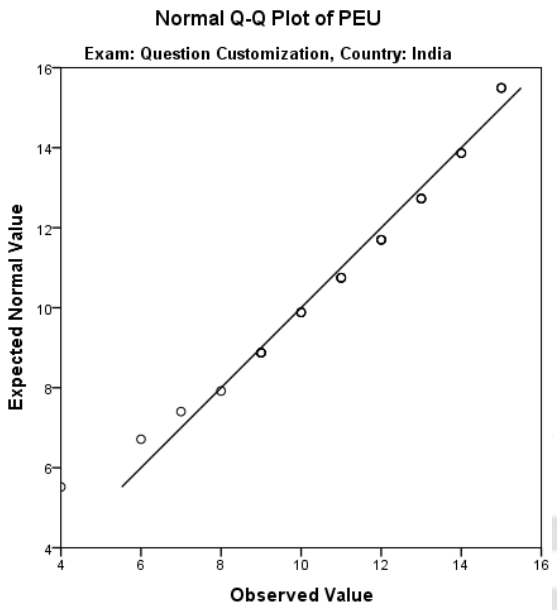


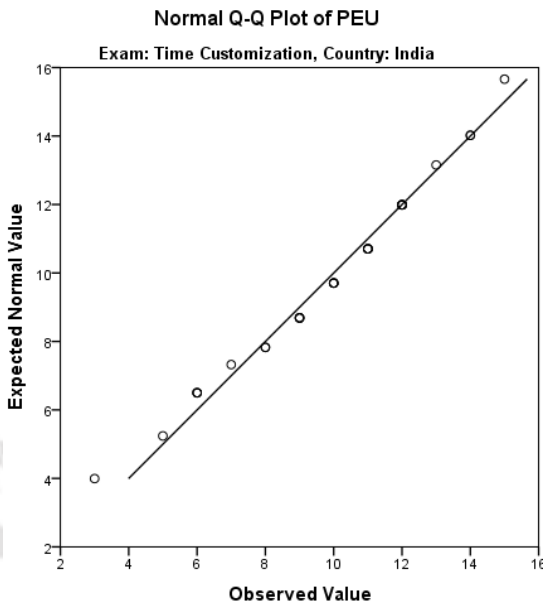


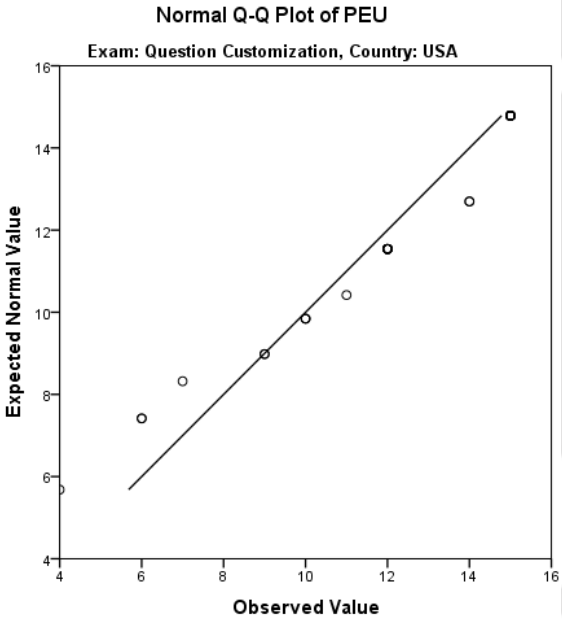


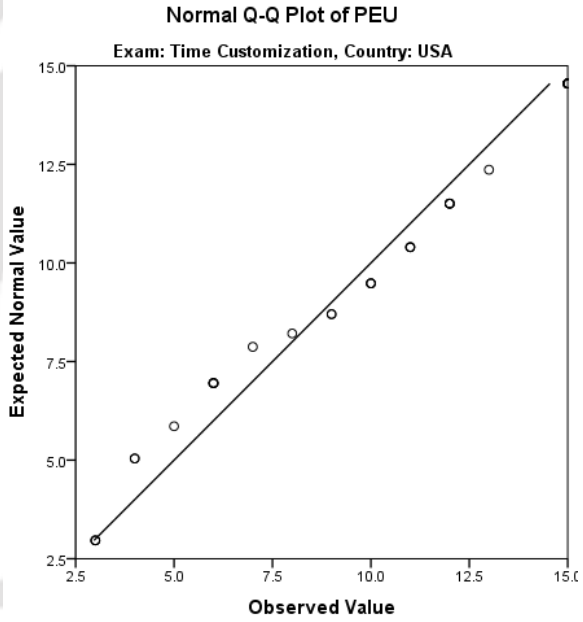


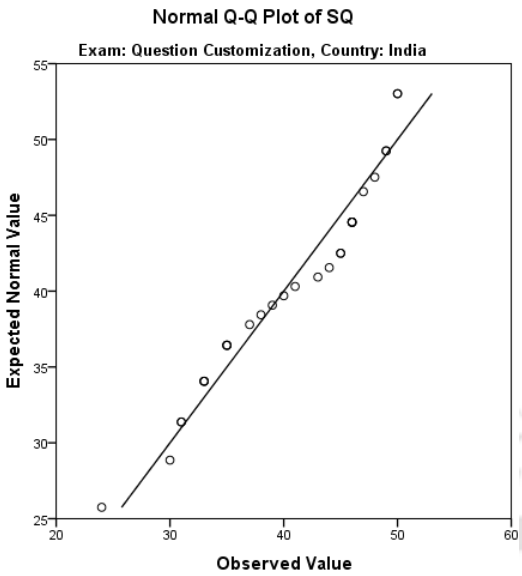


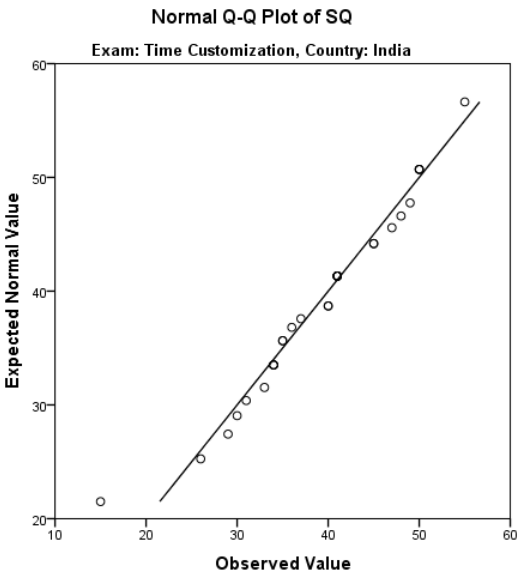


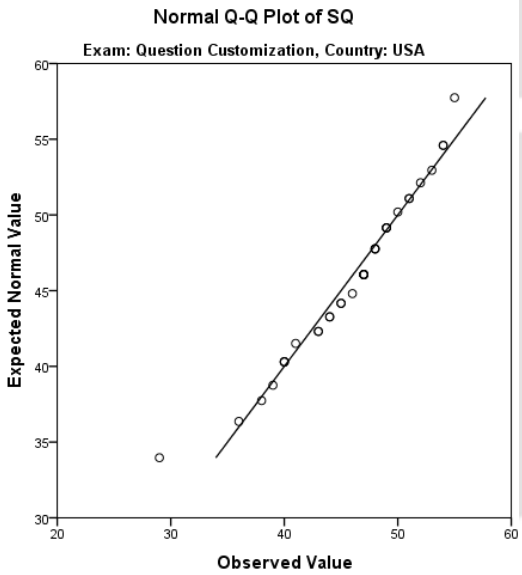


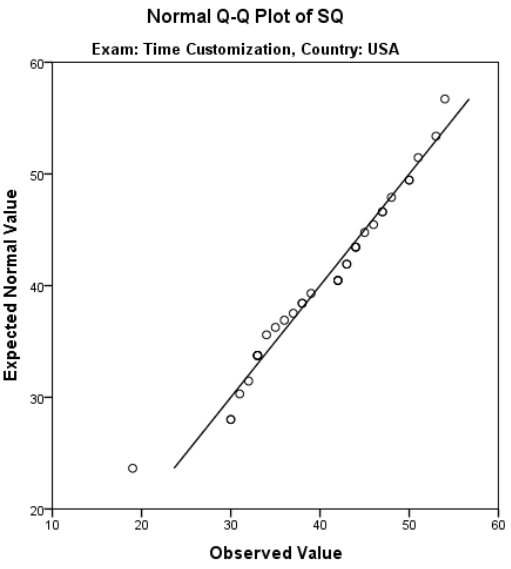


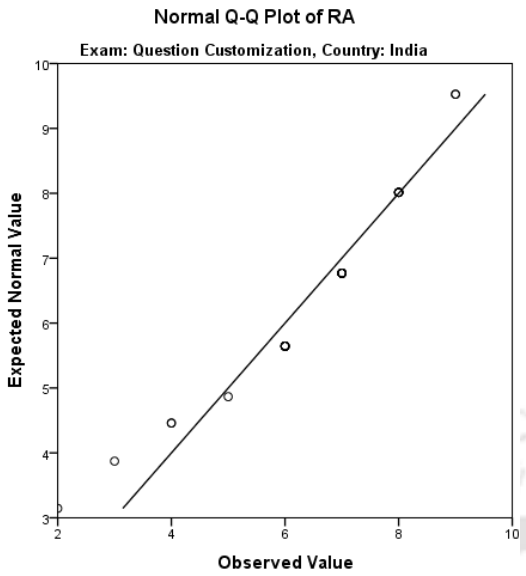


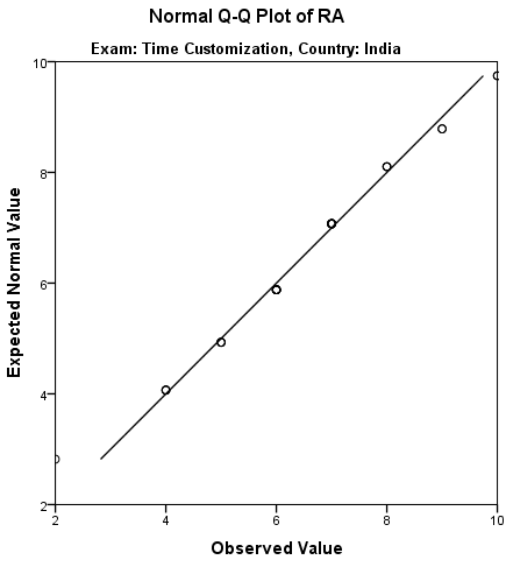


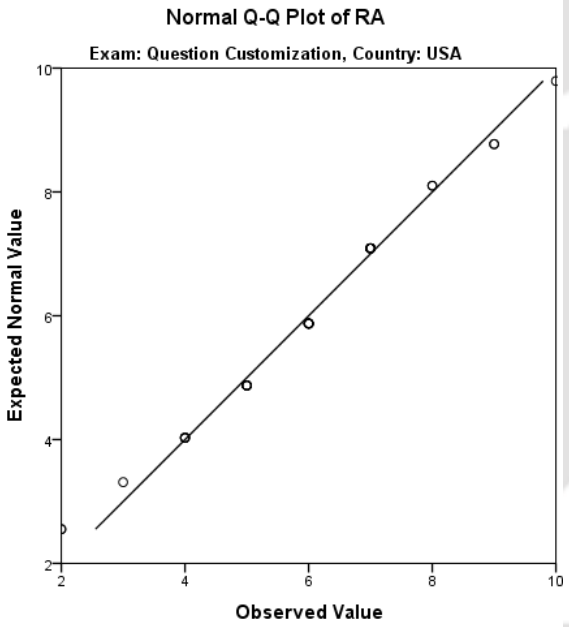


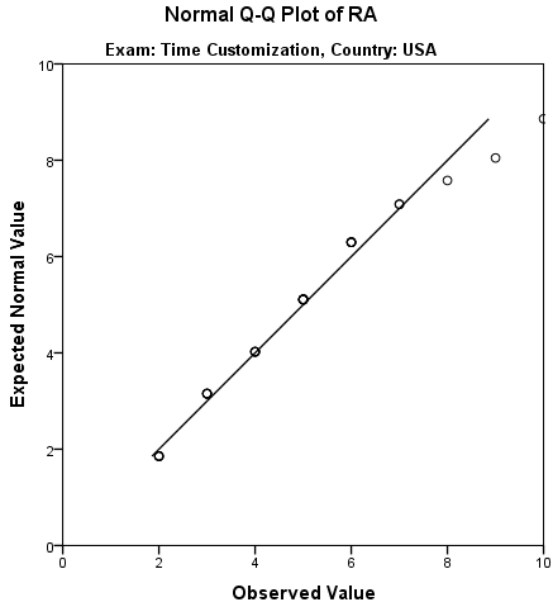


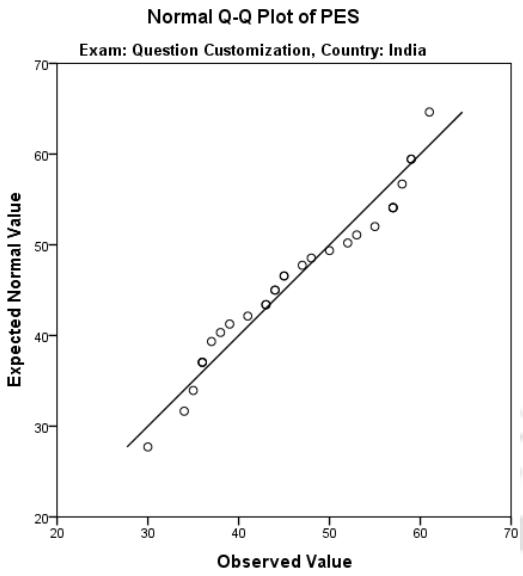


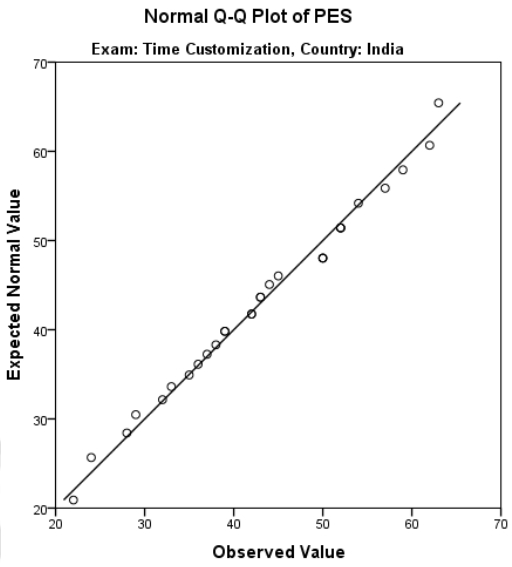


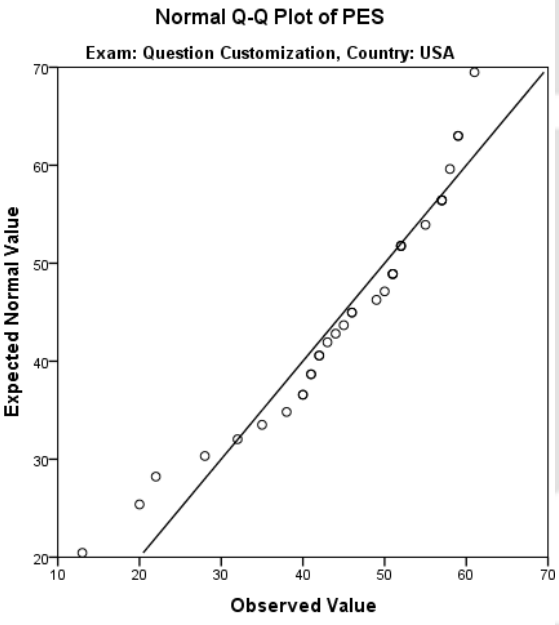


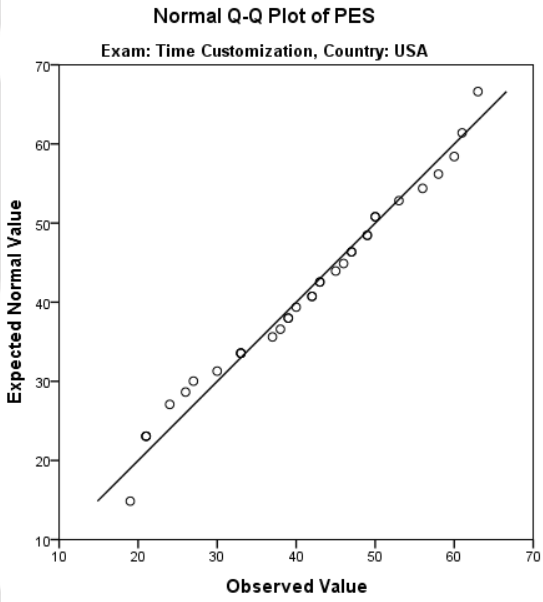


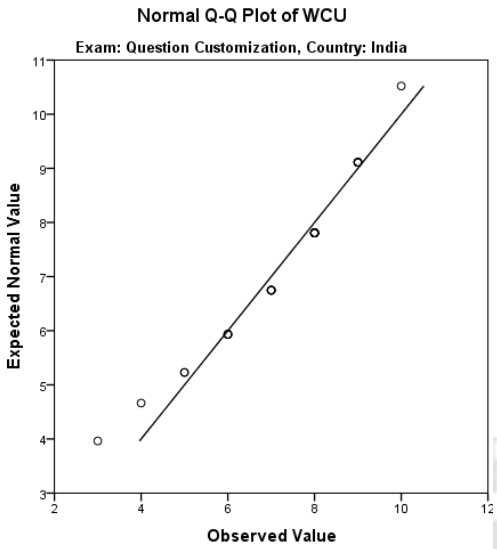


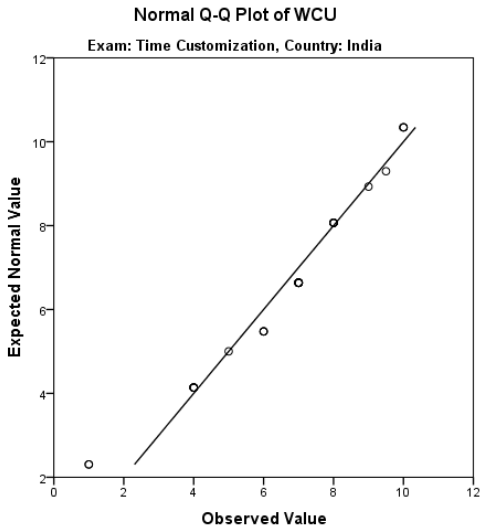


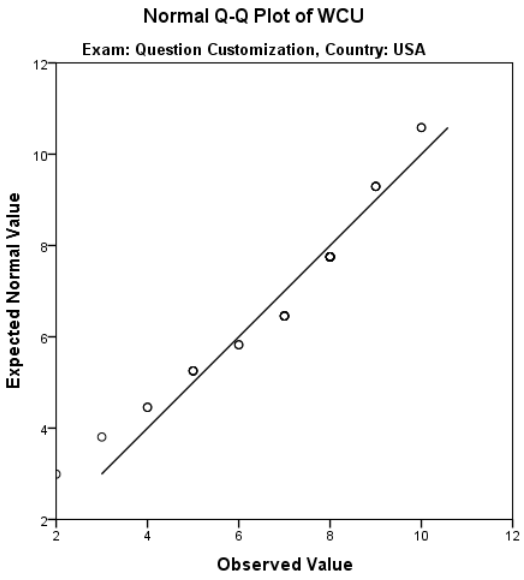


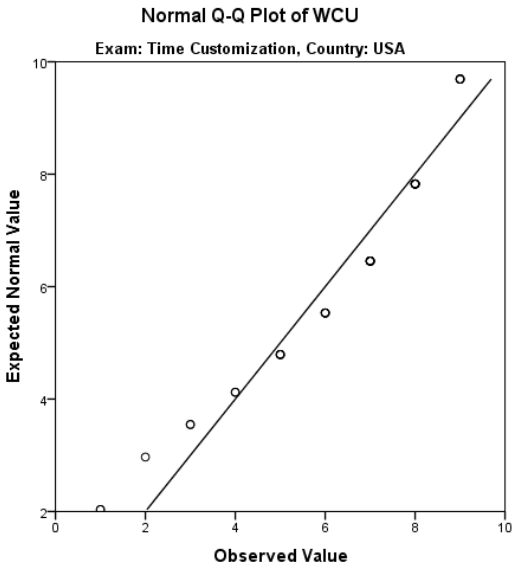


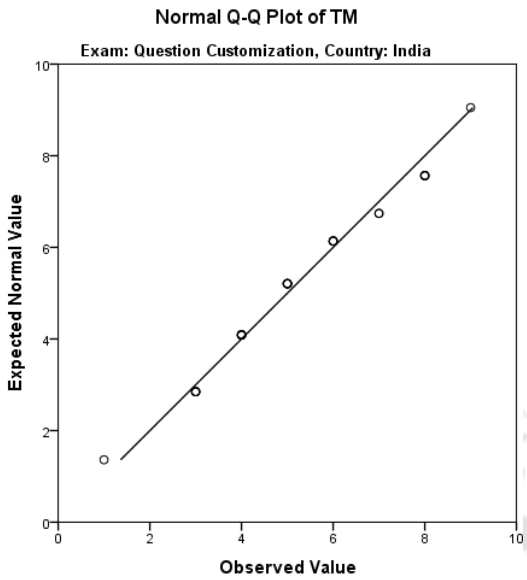


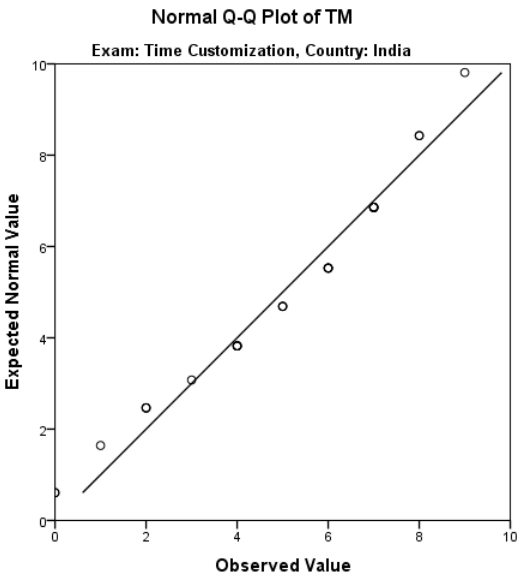


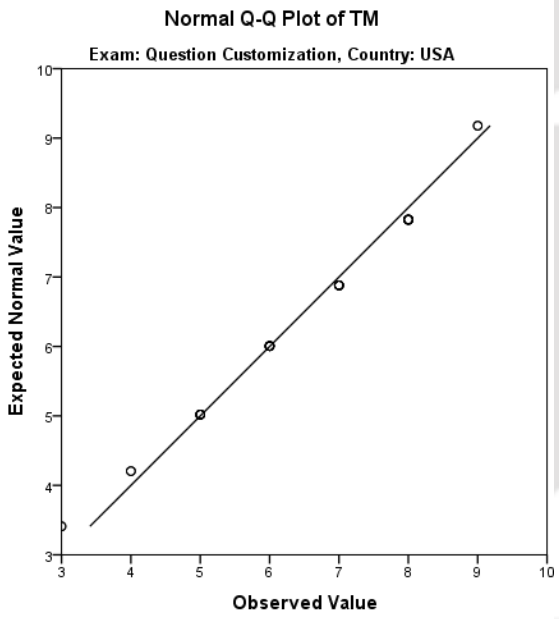


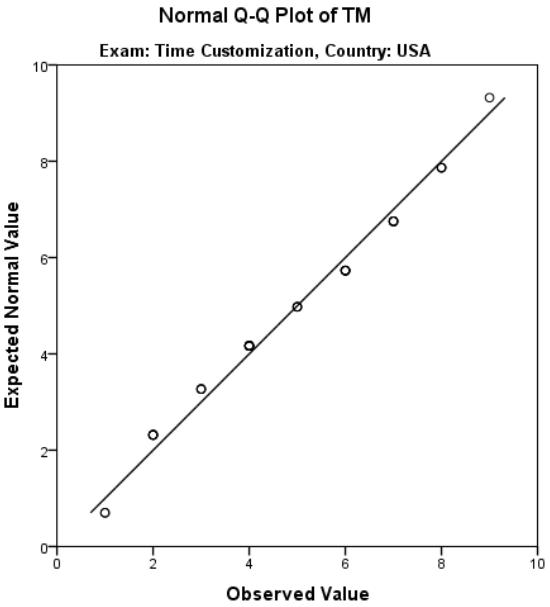


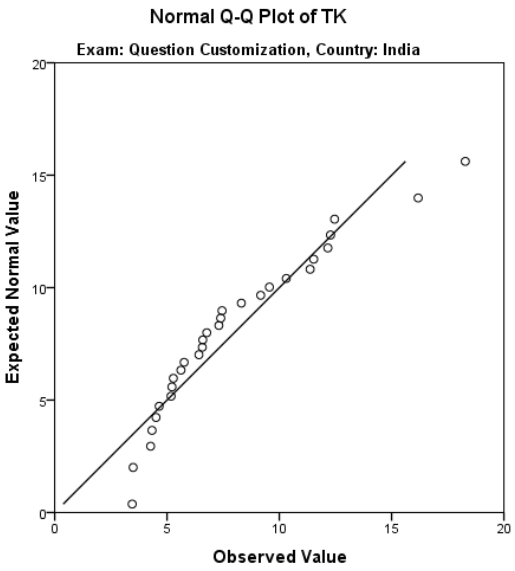


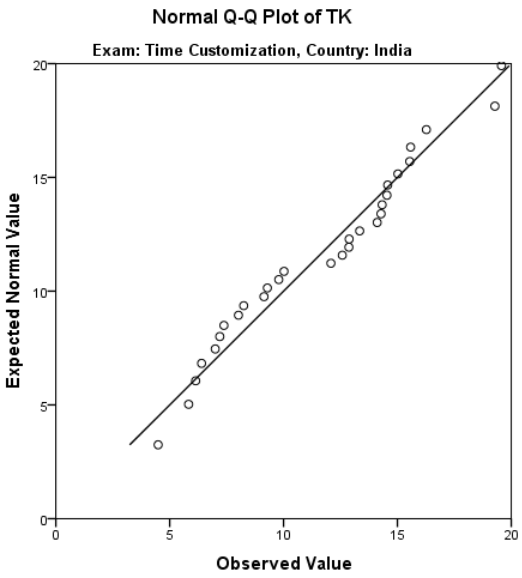


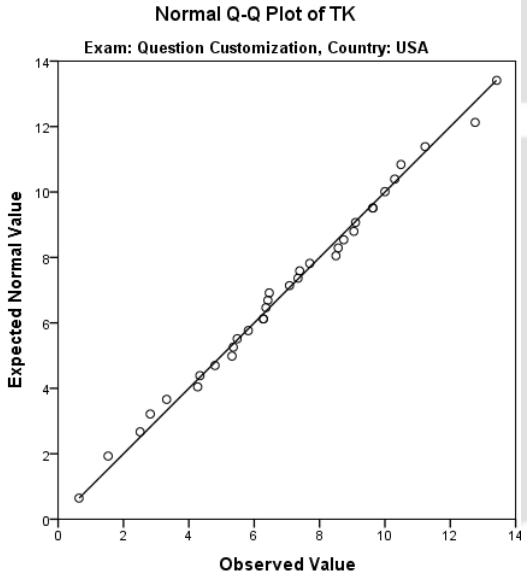












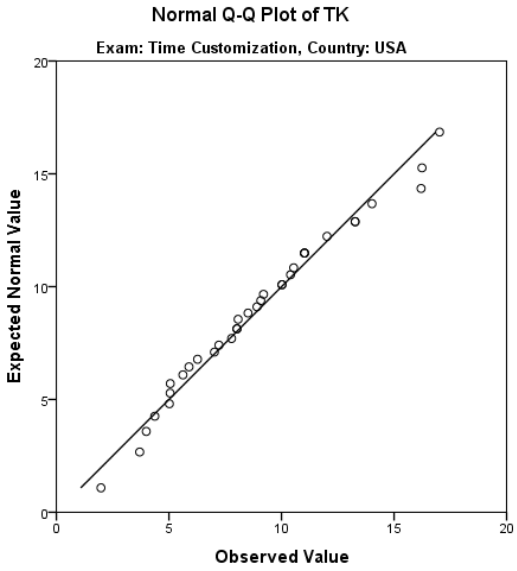


Table 3C.1

Box's test of equality of covariance matrices.

Box's M	280.36
F	1.44
df1	16
df2	30877.84
Sig.	.00

Box's test of equality of covariance matrices highlights that it is significant at $p<0.005$. Therefore, the assumption of homogeneity of variance-covariance of matrices has been violated. It suggests that the variance-covariance matrices among the dependent variables across the groups are not the same.

Table 3C.2

Levene's Test of Equality of Error Variances[a]

	F	df1	df2	Sig.
NASATLX	4.03	3	122	.00
SUS	1.64	3	122	.18
PEOU	5.36	3	122	.00
PEU	3.71	3	122	.01
SQ	2.32	3	122	.07
RA	.09	3	122	.96
PES	.65	3	122	.58
WCU	1.53	3	122	.20
TM	2.68	3	122	.05
TK	1.56	3	122	.20

Levene's test of equality of variances highlights that it is significant at $p<0.05$ for NASA TLX, PEOU, PEU; and not significant for SUS, SQ, RA, PES, WCU, TM and TK. Therefore, the assumption of homogeneity of variances assumption has been violated for NASA TLX, PEOU, PEU.

Table 3C.3

Univariate statistics of the dependent measures on the interacting Independent measures

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power[k]
	NASATLX	401.99	1	401.99	1.60	.21	.01	.24
	SUS	83.92	1	83.92	.31	.58	.00	.09
	PEOU	16.76	1	16.76	1.15	.29	.01	.19
	PEU	8.11	1	8.11	.74	.39	.01	.14
Exam *	**SQ**	153.93	1	153.93	2.81	.10	.02	.38
Country	**RA**	9.18	1	9.18	2.75	.10	.02	.38
	PES	11.94	1	11.94	.10	.76	.00	.06
	WCU	.11	1	.11	.02	.88	.00	.05
	TM	12.20	1	12.20	2.95	.09	.02	.40
	TK	21.17	1	21.17	1.57	.21	.01	.24

Table 3C.4(a)

Pairwise comparisons of the dependent measures between the Exam conditions and the countries.

Dependent Variable	Exam	Country	Mean	Std. Error	95% Confidence Interval	
					Lower Bound	Upper Bound
NASATLX	Question	India	34.78	2.94	28.96	40.60
	Customization	USA	34.42	2.71	29.05	39.80
	Time	India	39.96	2.94	34.14	45.78
	Customization	USA	46.77	2.71	41.40	52.15
SUS	Question	India	72.07	3.06	66.00	78.13
	Customization	USA	81.47	2.83	75.87	87.07
	Time	India	63.14	3.06	57.07	69.20
	Customization	USA	69.26	2.83	63.66	74.87
PEOU	Question	India	16.10	.71	14.70	17.51
	Customization	USA	16.88	.66	15.58	18.18
	Time	India	14.66	.71	13.25	16.06
	Customization	USA	13.97	.66	12.67	15.27
PEU	Question	India	11.10	.62	9.88	12.32
	Customization	USA	12.24	.57	11.11	13.36
	Time	India	9.83	.62	8.61	11.05
	Customization	USA	9.94	.57	8.81	11.07
SQ	Question	India	40.31	1.37	37.59	43.03
	Customization	USA	45.85	1.27	43.34	48.36
	Time	India	39.07	1.37	36.35	41.79
	Customization	USA	40.18	1.27	37.66	42.69

Table 3C.4 (b)

Pairwise comparisons of the dependent measures between the Exam conditions and the countries.

Dependent Variable	Exam	Country	Mean	Std. Error	95% Confidence Interval	
					Lower Bound	Upper Bound
RA	Question	India	6.55	.34	5.88	7.22
	Customization	USA	5.94	.31	5.32	6.56
	Time	India	6.52	.34	5.85	7.19
	Customization	USA	4.82	.31	4.20	5.44
PES	Question	India	46.17	2.08	42.06	50.29
	Customization	USA	44.97	1.92	41.17	48.77
	Time	India	43.17	2.08	39.06	47.29
	Customization	USA	40.74	1.92	36.93	44.54
WCU	Question	India	7.24	.39	6.48	8.00
	Customization	USA	7.03	.36	6.32	7.73
	Time	India	6.53	.39	5.77	7.30
	Customization	USA	6.21	.36	5.50	6.91
TM	Question	India	5.21	.38	4.46	5.95
	Customization	USA	6.29	.35	5.60	6.98
	Time	India	4.90	.38	4.15	5.64
	Customization	USA	4.74	.35	4.04	5.43
TK	Question	India	8.00	.68	6.65	9.35
	Customization	USA	7.03	.63	5.78	8.27
	Time	India	11.58	.68	10.23	12.93
	Customization	USA	8.97	.63	7.72	10.21

Table 3C.5

Correlation of the dependent variables

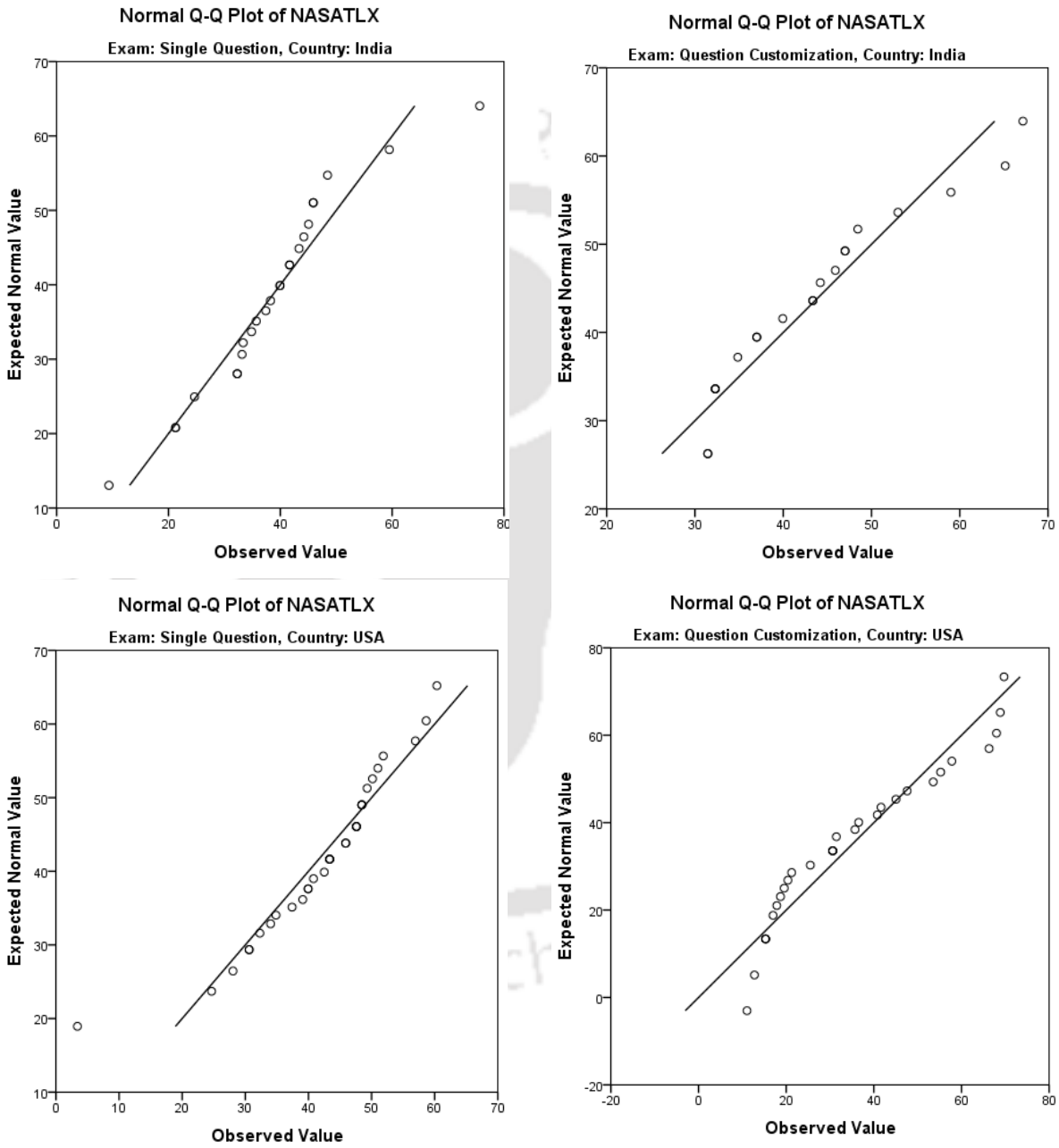
	1	2	3	4	5	6	7	8	9	10
1. NASATLX	1.00									
2. SUS	-.37**	1.00								
3. PEOU	-.46**	.69**	1.00							
4. PEU	-.45**	.60**	.72**	1.00						
5. SQ	-.31**	.65**	.65**	.60**	1.00					
6. RA	-.17	.31**	.43**	.33**	.34**	1.00				
7. PES	-.38**	.65**	.69**	.64**	.65**	.54**	1.00			
8. WCU	-.13	.39**	.52**	.49**	.50**	.58**	.67**	1.00		
9. TM	-.38**	.46**	.35**	.37**	.41**	.27**	.33**	.23*	1.00	
10. TK	.08	-.22*	-.21*	-.20*	-.20*	.05	-.14	-.12	-.06	1.00

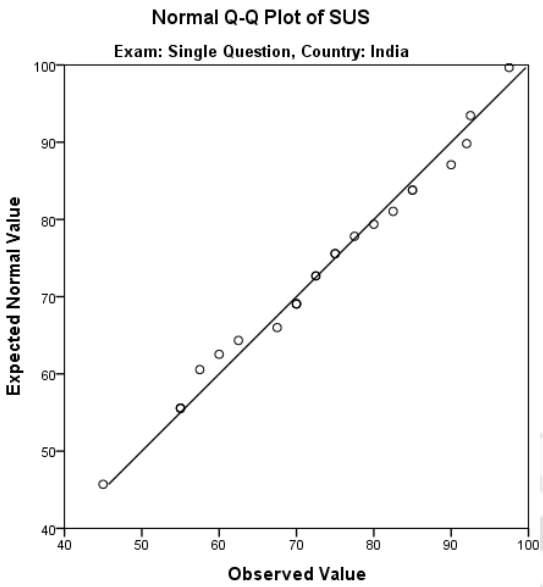
**p<0.01 (2-tailed), *p<0.05 (2-tailed).

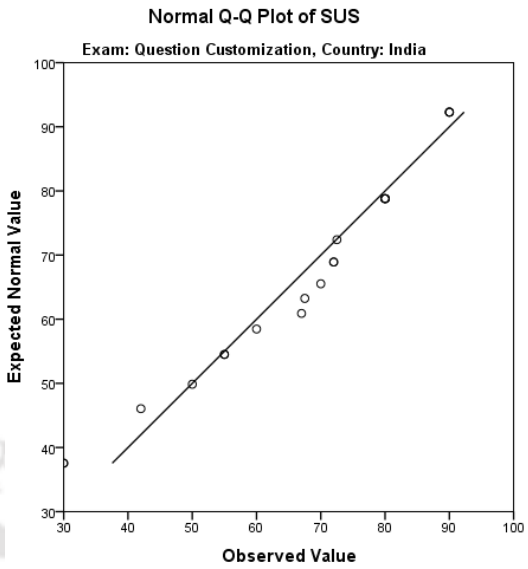
Appendix – 3D

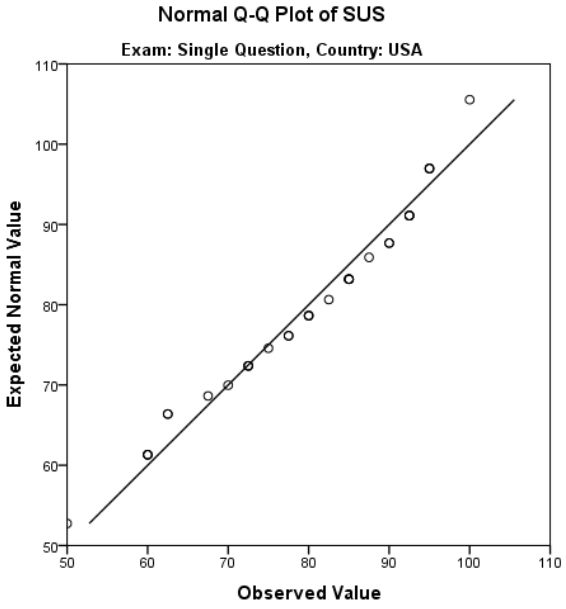
Experiment 4: Single Question without Customization vs. Question Customization

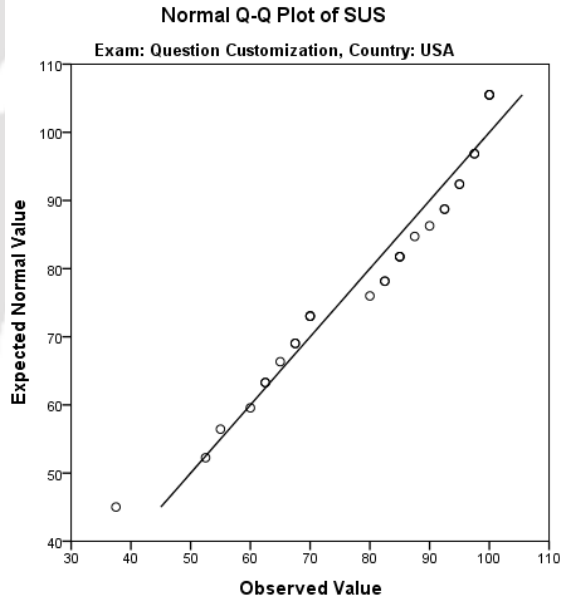
Figure 3D.1 Normal Q-Q plots

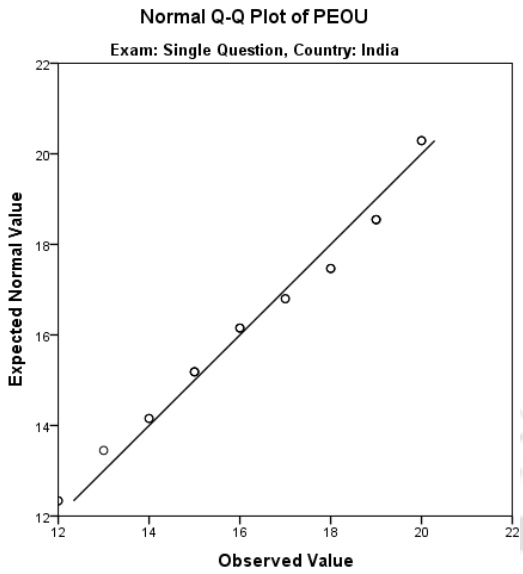


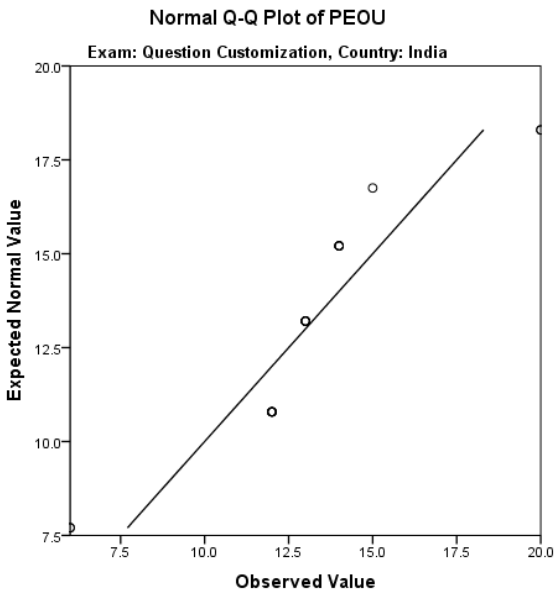


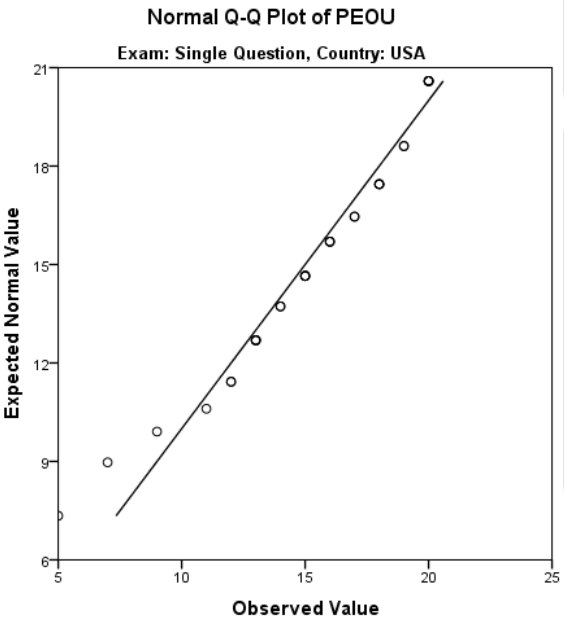


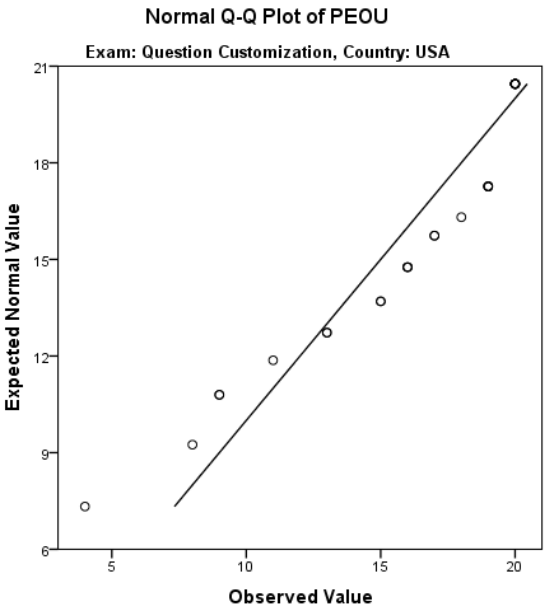


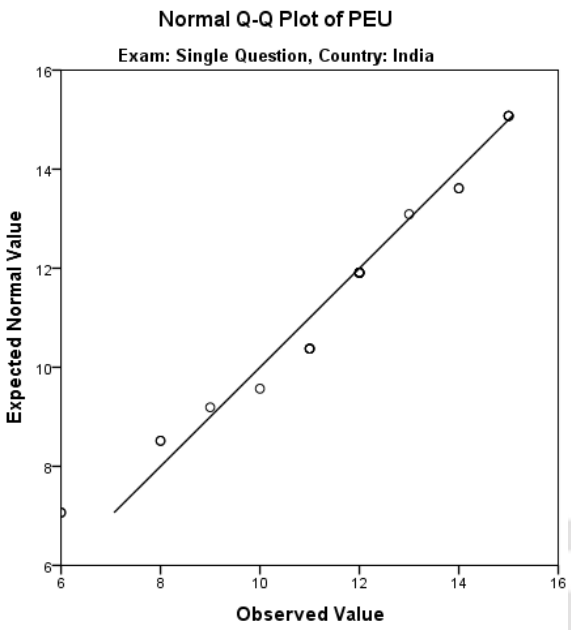


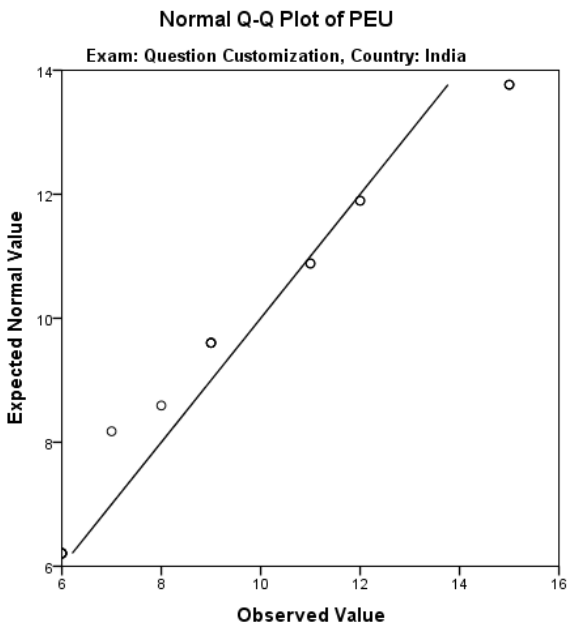


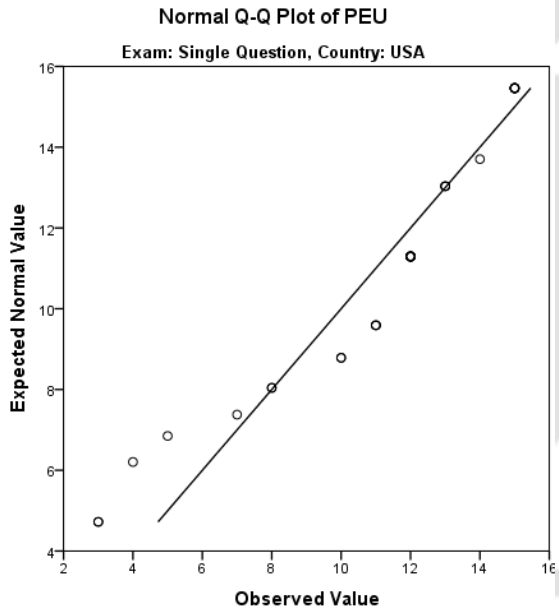


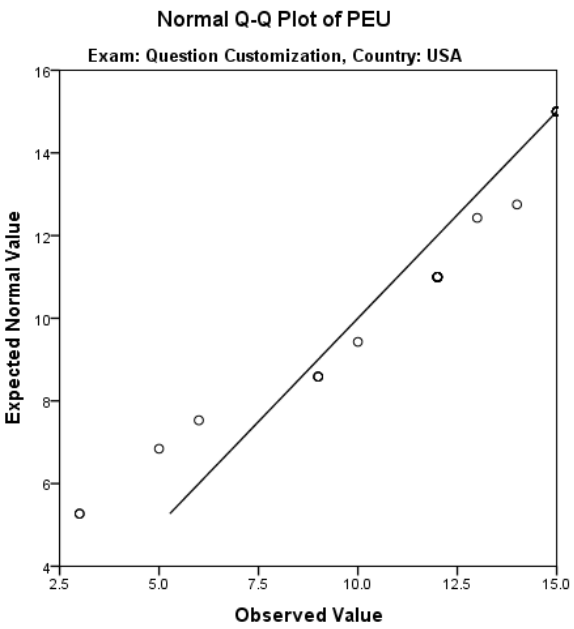


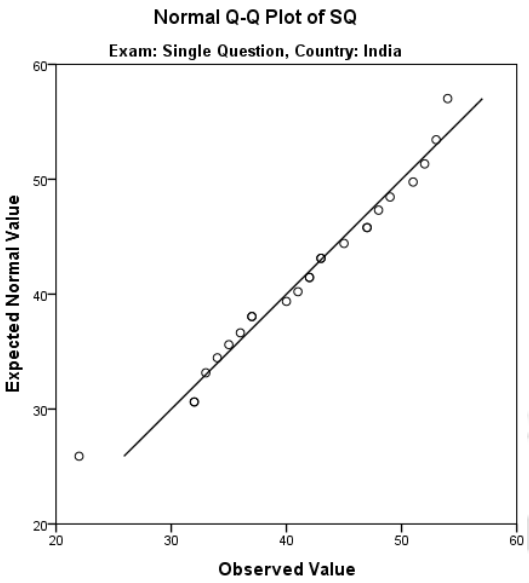


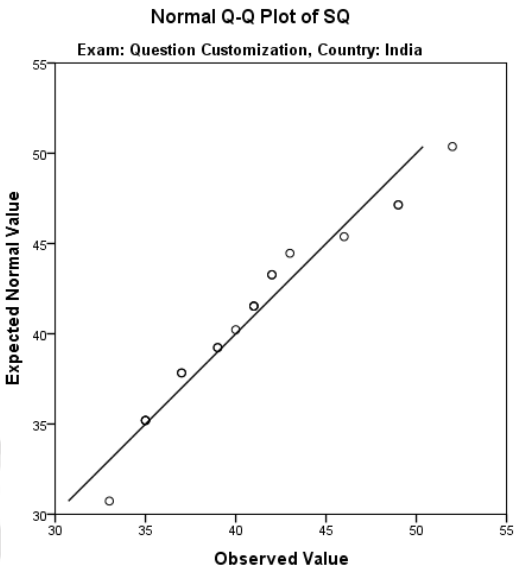


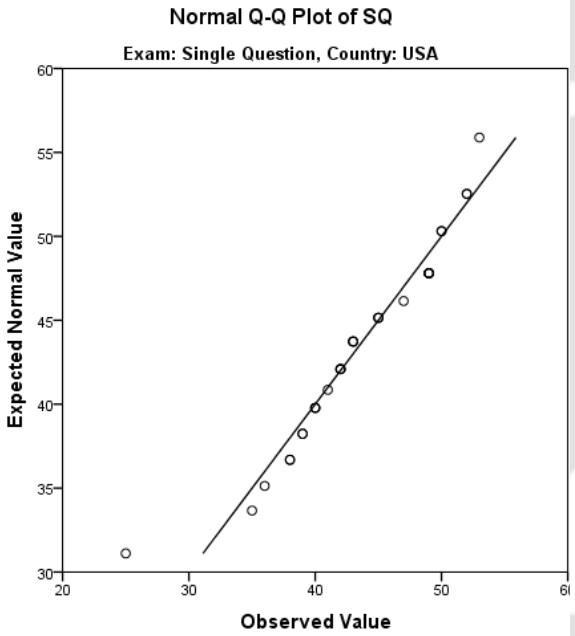


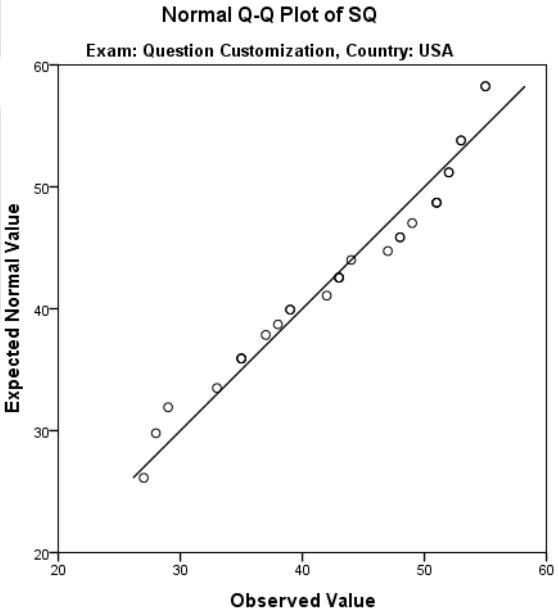


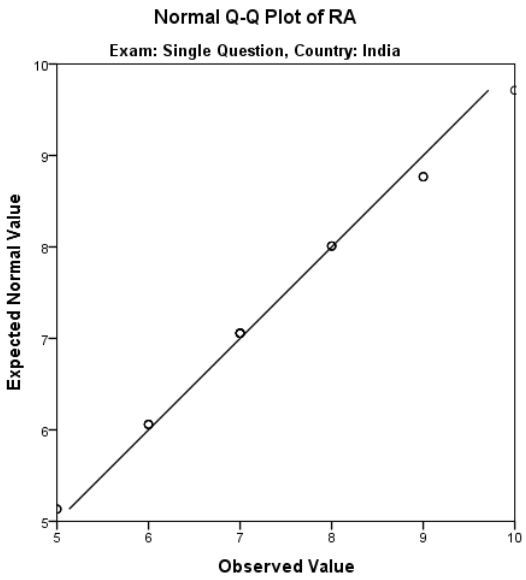


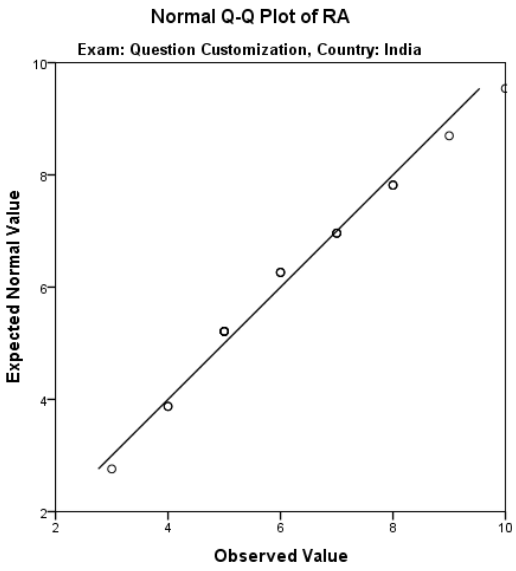


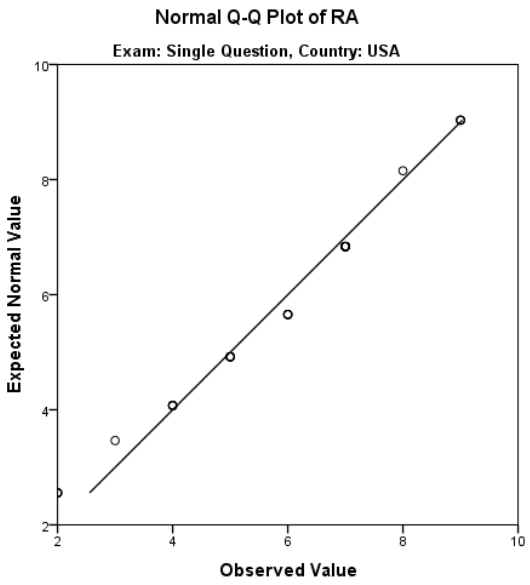


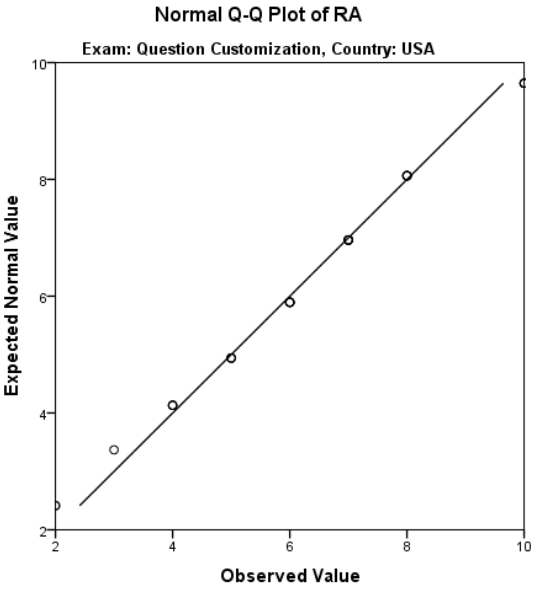


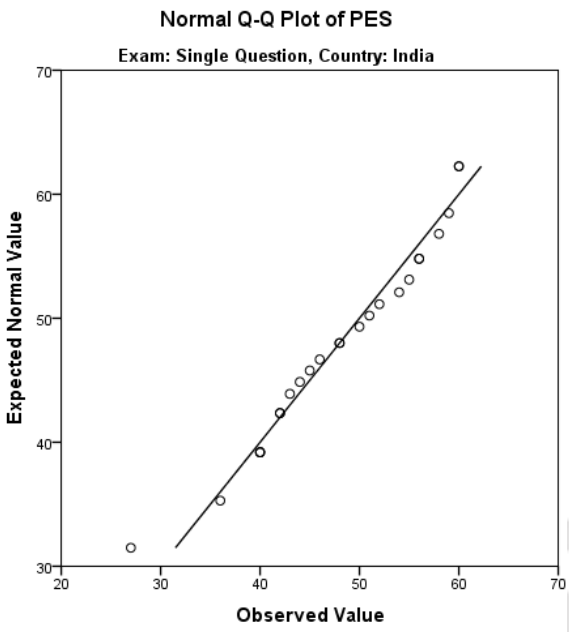


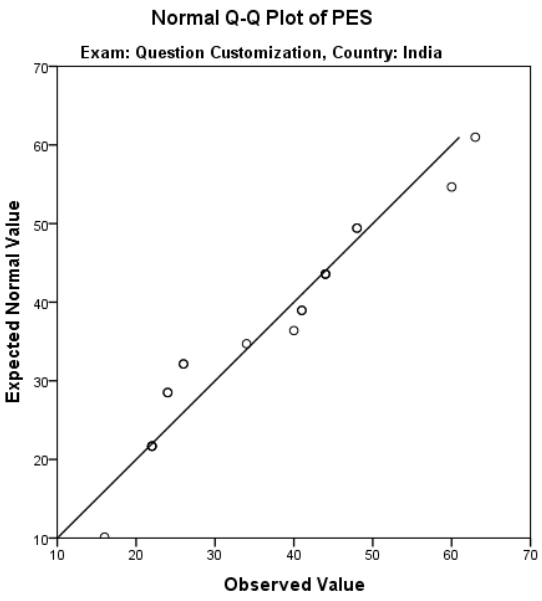


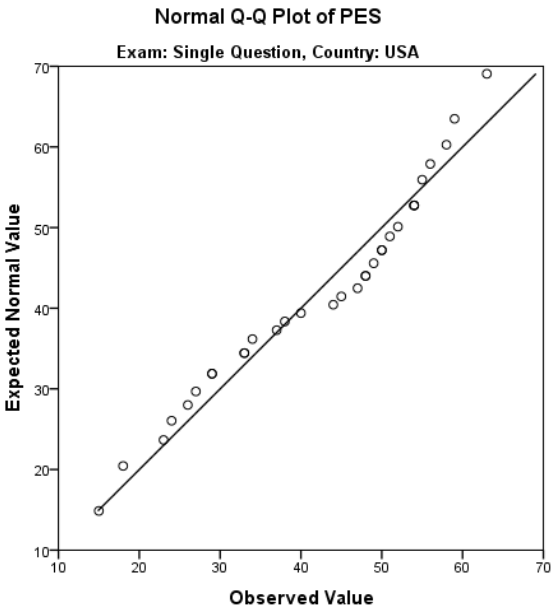


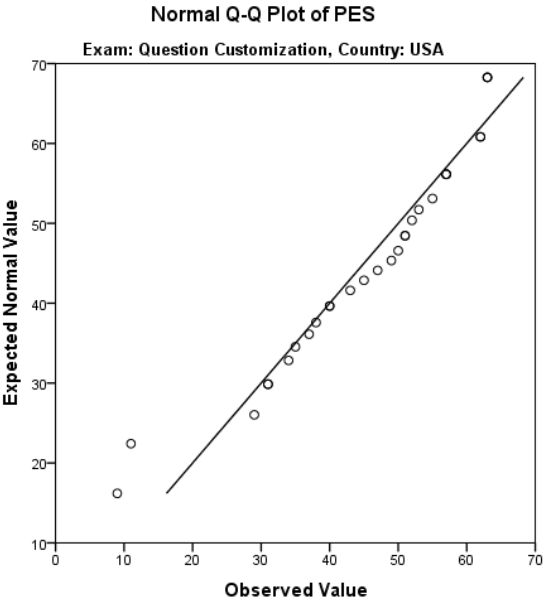


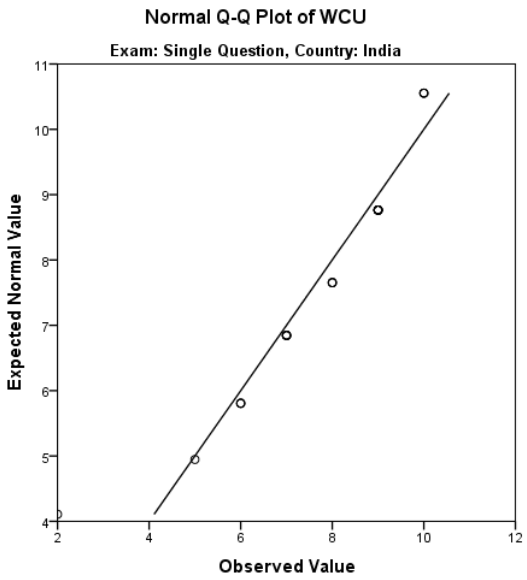


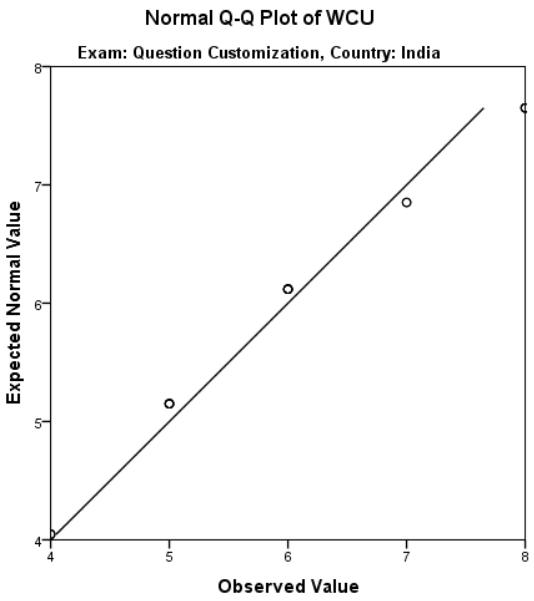


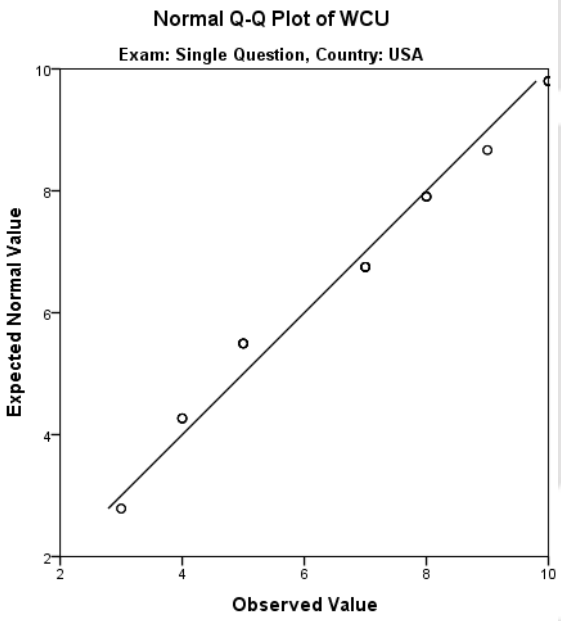


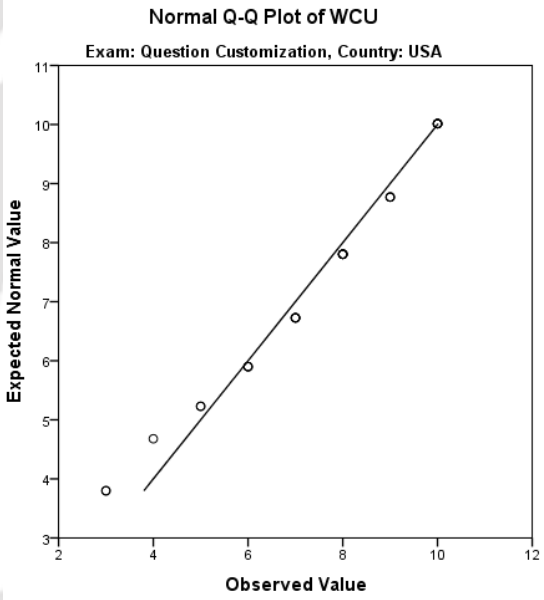


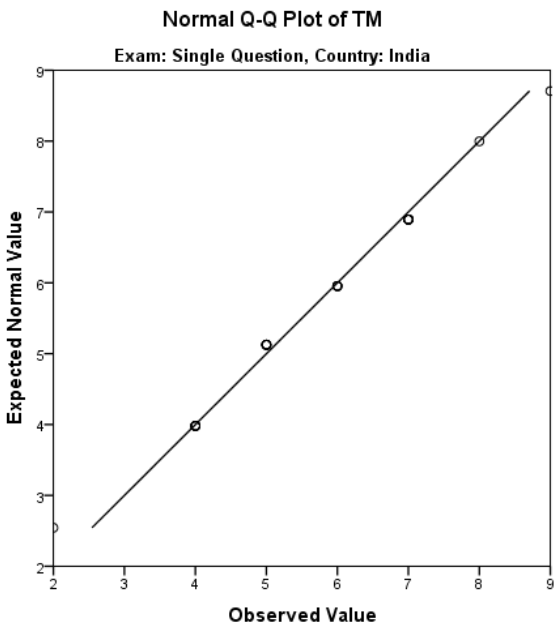


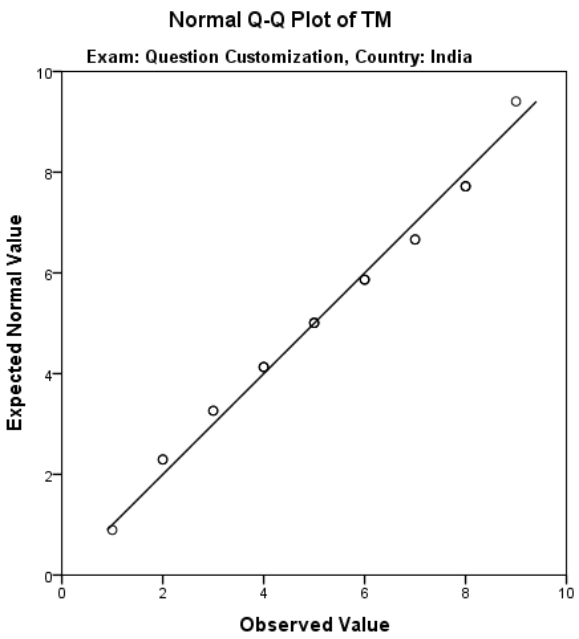


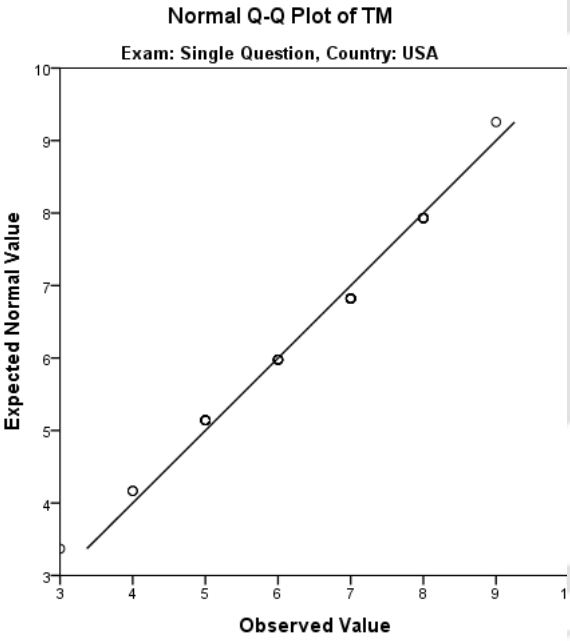


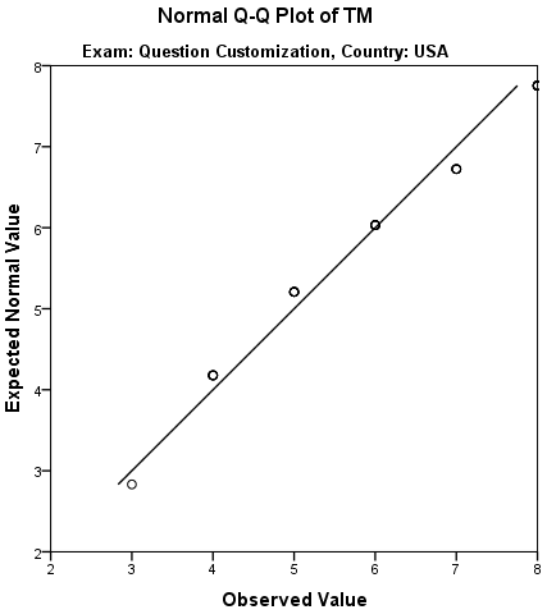


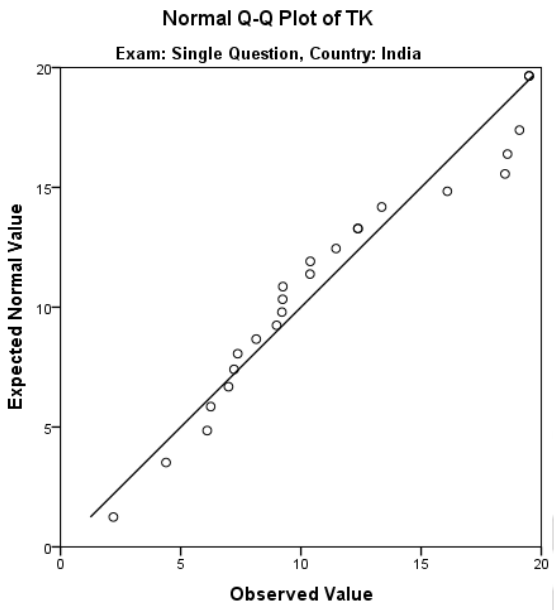


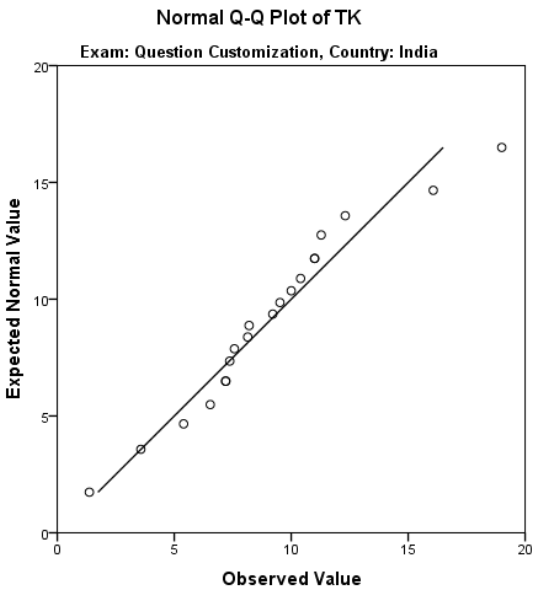


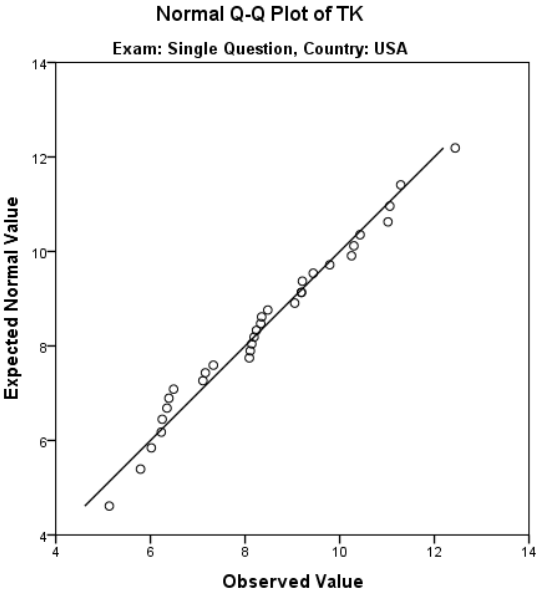












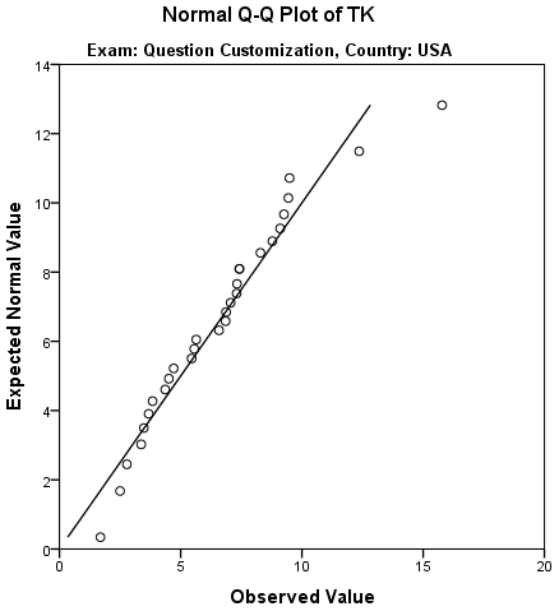


Table 3D.1

Box's test of equality of covariance matrices.

Box's M	335.09
F	1.65
df1	165
df2	17629.48
Sig.	.00

Box's test of equality of covariance matrices highlights that it is significant at $p<0.005$. Therefore, the assumption of homogeneity of variance-covariance of matrices has been violated. It suggests that the variance-covariance matrices among the dependent variables across the groups are not the same.

Table 3D.2

Levene's Test of Equality of Error Variances[a]

	F	**df1**	**df2**	**Sig.**
NASATLX	5.26	3	101	.00
SUS	1.62	3	101	.18
PEOU	2.48	3	101	.06
PEU	.73	3	101	.53
SQ	3.25	3	101	.02
RA	2.54	3	101	.06
PES	2.95	3	101	.03
WCU	1.99	3	101	.11
TM	2.55	3	101	.06
TK	7.41	3	101	.00

Levene's test of equality of variances highlights that it is significant at $p<0.05$ for NASA TLX, SQ, PES, TK and not significant for SUS, PEOU, PEU, RA, WCU and TM. Therefore, the assumption of homogeneity of variances assumption has been violated for NASA TLX, SQ, PES and TK.

Table 3D.3

Univariate statistics of the dependent measures based on the independent factor - Exam.

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power[k]
Exam	**NASATLX**	21.43	1.00	21.43	.11	.74	.00	.06
	SUS	252.07	1.00	252.07	1.08	.30	.01	.18
	PEOU	39.12	1.00	39.12	2.91	.09	.03	.39
	PEU	17.57	1.00	17.57	1.57	.21	.02	.24
	SQ	8.15	1.00	8.15	.16	.69	.00	.07
	RA	.40	1.00	.40	.11	.74	.00	.06
	PES	523.10	1.00	523.10	3.24	.07	.03	.43
	WCU	7.71	1.00	7.71	2.04	.16	.02	.29
	TM	4.29	1.00	4.29	1.52	.22	.01	.23
	TK	92.84	1.00	92.84	7.45	.01	.07	.77

Table 3D.4

Independent sample t-test across the interface features among Indian and American students

Single Question without customization	Levene's Test for Equality of Variances				
	F	Sig.	t	df	Sig. (2-tailed)
SUS	.07	.79	-1.80	54.00	.08
RA	6.53	.01	3.05	54.00	.00
PES	8.47	.01	1.96	54.00	.05
WCU	1.28	.26	1.80	54.00	.08
TM	.23	.64	-1.70	54.00	.09
TK	24.15	.00	2.81	54.00	.01
Question Customization					
NASATLX	8.66	.01	1.80	47.00	.08
SUS	.01	.94	-2.42	47.00	.02
PEOU	3.84	.06	-2.83	47.00	.01
PEU	.04	.85	-2.54	47.00	.01
PES	.01	.90	-2.39	47.00	.02
WCU	3.55	.07	-2.85	47.00	.01
TK	.62	.43	2.52	47.00	.02

Table 3D.5

Independent sample t-test across the Indian and American students among the CBA conditions

India	Levene's Test for Equality of Variances				
	F	Sig.	*t*	*df*	Sig. (2-tailed)
PEOU	.04	.84	4.03	42.00	.00
PEU	1.61	.21	2.59	42.00	.01
RA	2.98	.09	2.05	42.00	.05
PES	8.44	.01	3.70	42.00	.00
WCU	1.73	.20	3.59	42.00	.00
USA					
NASATLX	11.83	.00	1.76	59.00	.08
TK	4.87	.03	2.83	59.00	.01

Table 3D.6 (a)

Pairwise comparisons across the Indian and American students among the CBA conditions

Dependent Variable	Exam	Country	Mean	Std. Error	95% Confidence Interval	
					Lower Bound	Upper Bound
NASATLX	Single Question	India	38.54	2.87	32.84	44.24
		USA	42.07	2.49	37.14	47.01
	Question Customization	India	43.61	3.15	37.36	49.85
		USA	35.17	2.62	29.98	40.36
SUS	Single Question	India	72.69	3.12	66.49	78.88
		USA	79.14	2.70	73.78	84.50
	Question Customization	India	66.65	3.42	59.86	73.44
		USA	78.88	2.84	73.24	84.51
PEOU	Single Question	India	16.54	.75	15.06	18.03
		USA	15.25	.65	13.96	16.54
	Question Customization	India	13.00	.82	11.37	14.63
		USA	16.31	.68	14.96	17.66
PEU	Single Question	India	11.50	.68	10.14	12.86
		USA	11.16	.59	9.98	12.33
	Question Customization	India	9.20	.75	7.71	10.69
		USA	11.79	.62	10.56	13.03
SQ	Single Question	India	41.46	1.45	38.58	44.34
		USA	43.50	1.26	41.00	46.00
	Question Customization	India	40.55	1.59	37.39	43.71
		USA	43.28	1.32	40.65	45.90

Table 3D.6 (b)

Pairwise comparisons across the Indian and American students among the CBA conditions

Dependent Variable	Exam	Country	Mean	Std. Error	95% Confidence Interval	
					Lower Bound	Upper Bound
RA	Single Question	India	7.13	.39	6.34	7.91
		USA	5.66	.34	4.98	6.33
	Question Customization	India	6.15	.43	5.30	7.00
		USA	6.38	.36	5.67	7.09
PES	Single Question	India	48.00	2.59	42.86	53.14
		USA	41.97	2.25	37.51	46.42
	Question Customization	India	35.55	2.84	29.91	41.19
		USA	45.34	2.36	40.66	50.03
WCU	Single Question	India	7.75	.40	6.96	8.54
		USA	6.75	.34	6.07	7.43
	Question Customization	India	5.95	.44	5.09	6.81
		USA	7.45	.36	6.73	8.17
TM	Single Question	India	5.63	.34	4.94	6.31
		USA	6.31	.30	5.72	6.90
	Question Customization	India	5.15	.38	4.40	5.90
		USA	5.97	.31	5.35	6.59
TK	Single Question	India	11.12	.72	9.69	12.55
		USA	8.40	.62	7.16	9.64
	Question Customization	India	9.12	.79	7.55	10.68
		USA	6.58	.66	5.28	7.88

Table 3D.7

Correlational matrix among the dependent measures

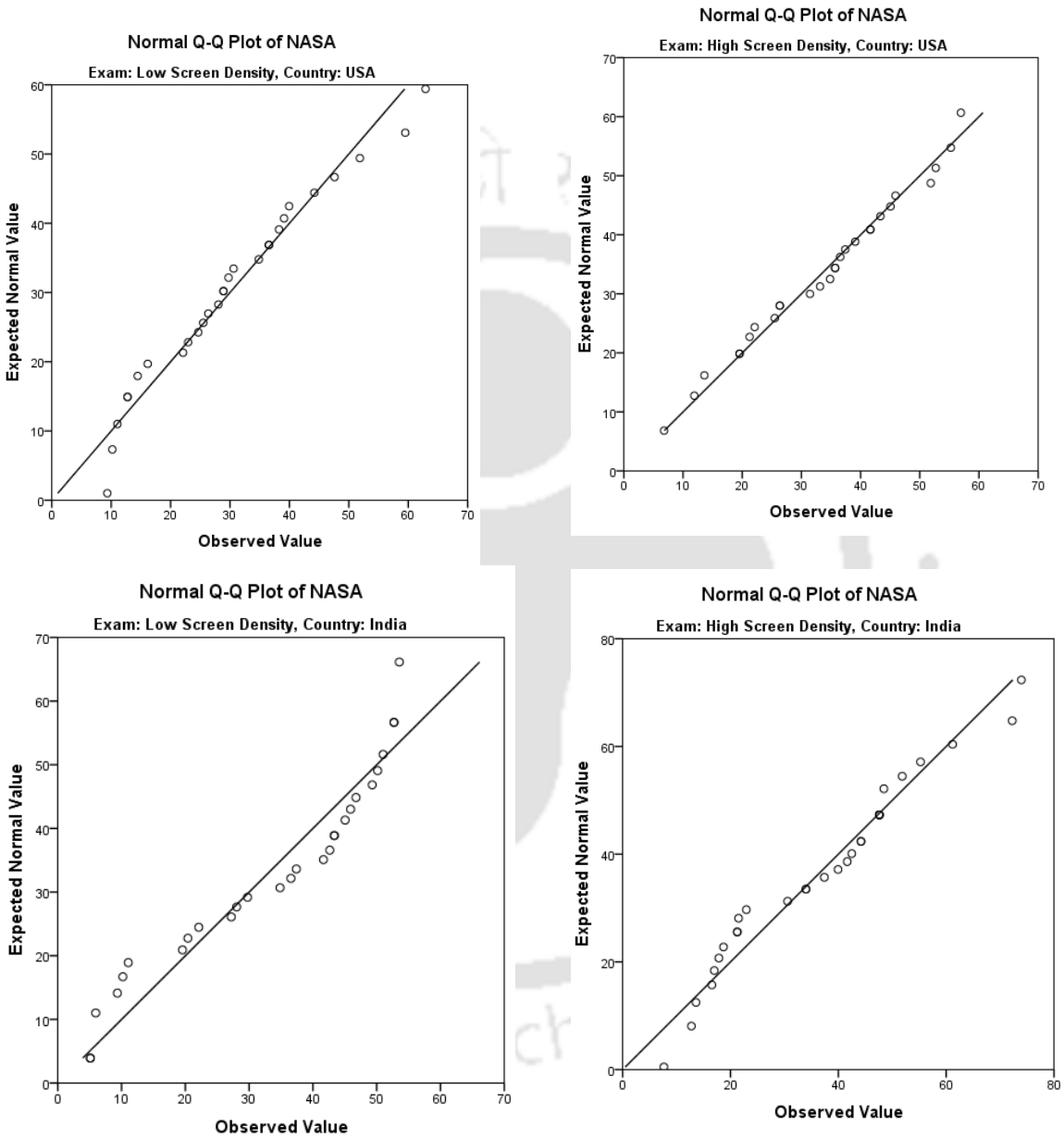
	1	2	3	4	5	6	7	8	9	10
1. NASATLX	1									
2. SUS	-.231*	1								
3. PEOU	-.417**	.559**	1							
4. PEU	-.411**	.606**	.823**	1						
5. SQ	-.172	.584**	.577**	.628**	1					
6. RA	-.248*	.346**	.495**	.531**	.479**	1				
7. PES	-.323**	.578**	.741**	.804**	.628**	.666**	1			
8. WCU	-.152	.421**	.537**	.470**	.380**	.574**	.599**	1		
9. TM	-.057	.114	.094	.160	.105	-.055	.115	.068	1	
10. TK	-.055	-.085	-.035	-.113	-.011	.102	-.019	.047	.071	1

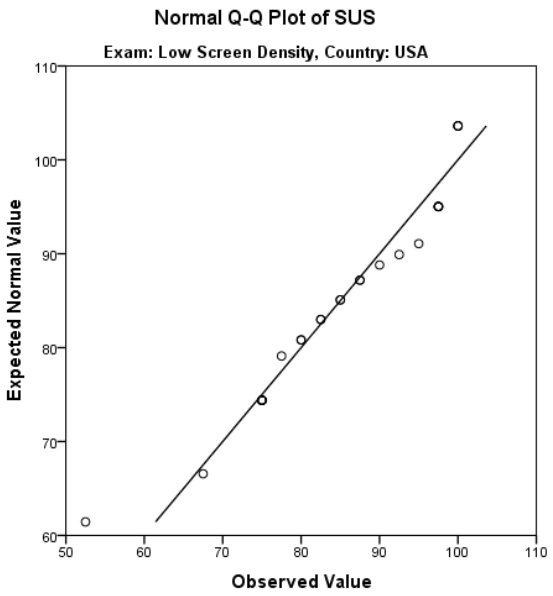
**p<0.01 (2-tailed), *p<0.05 (2-tailed).

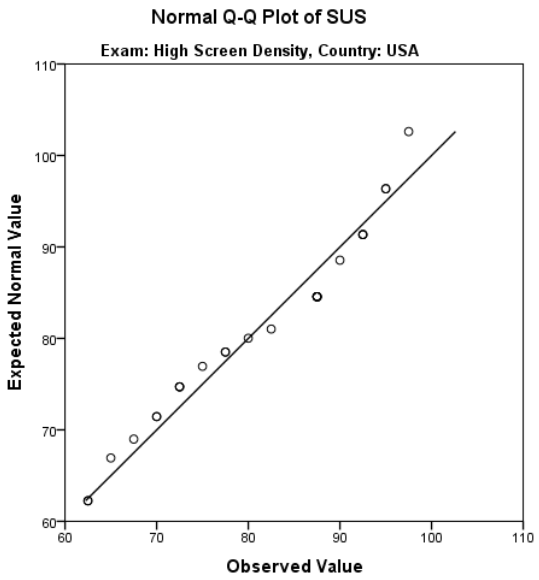
Appendix – 3E

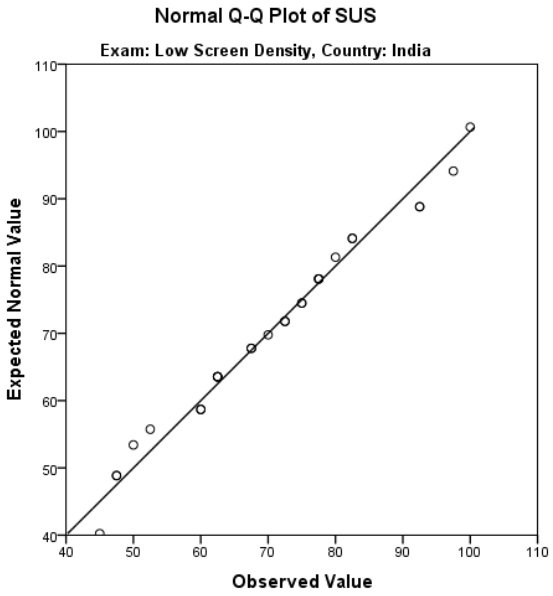
Experiment 5: Low Screen density vs. High Screen Density

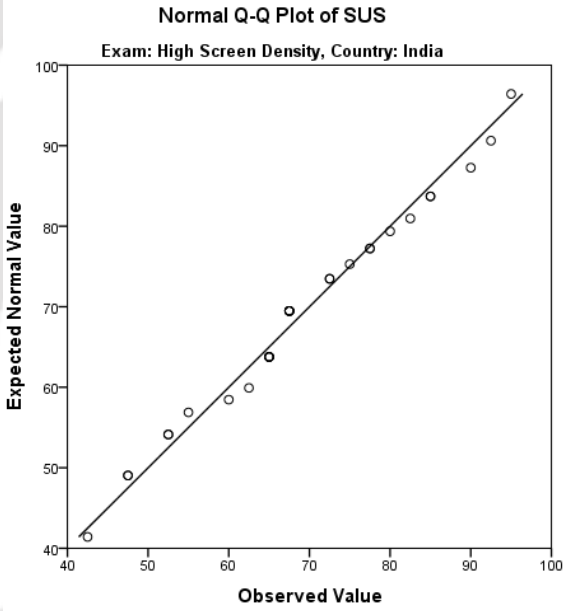
Figure 3E.1 Normal Q-Q plots

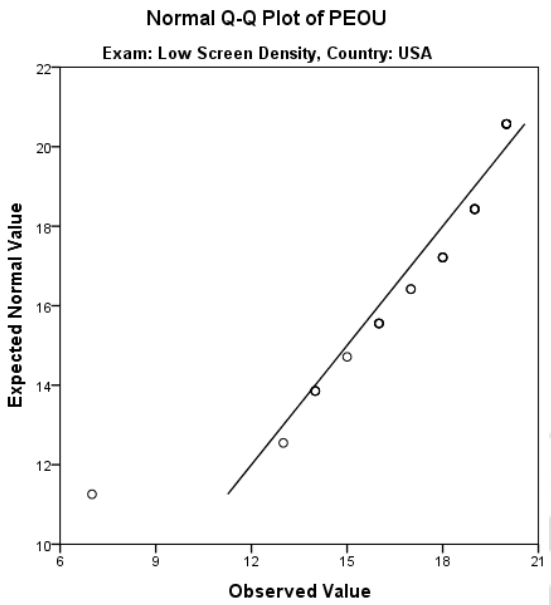


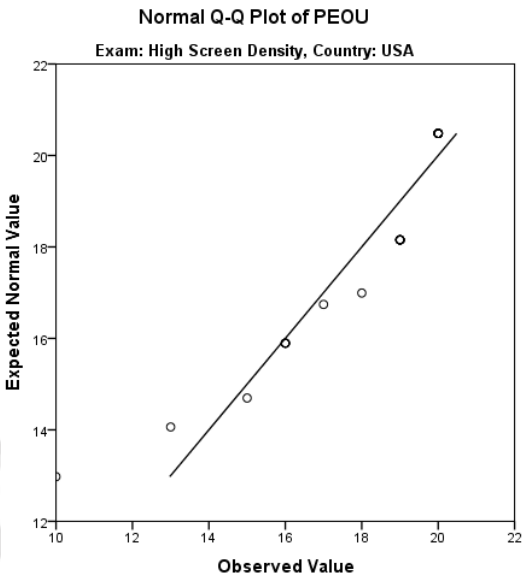


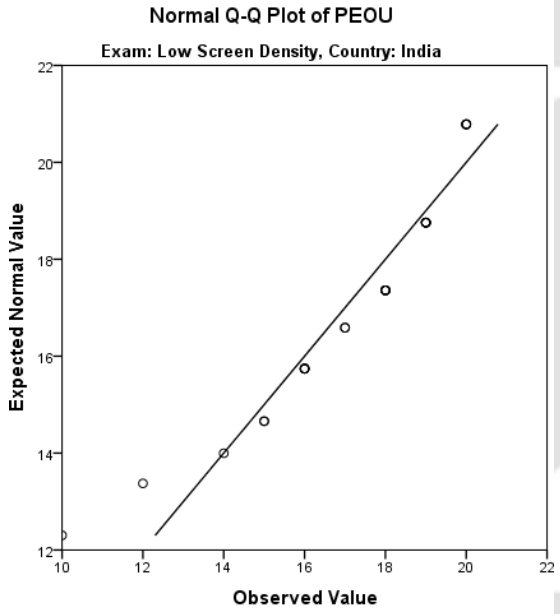


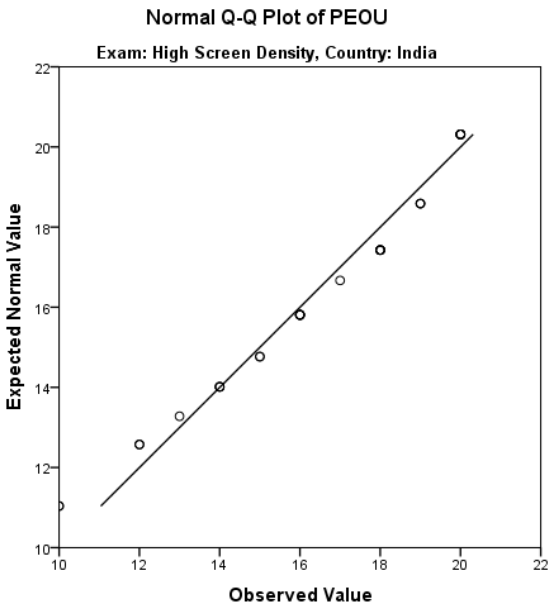


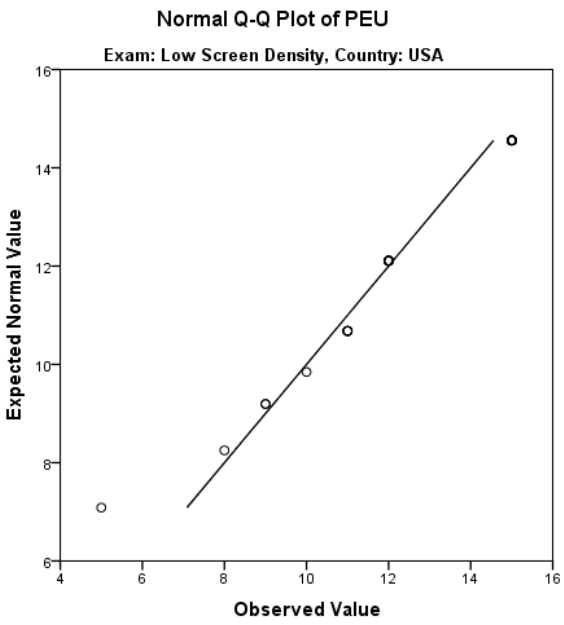


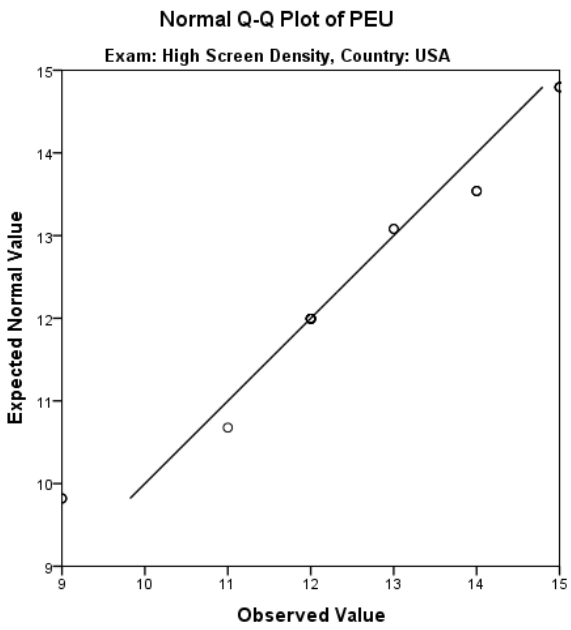


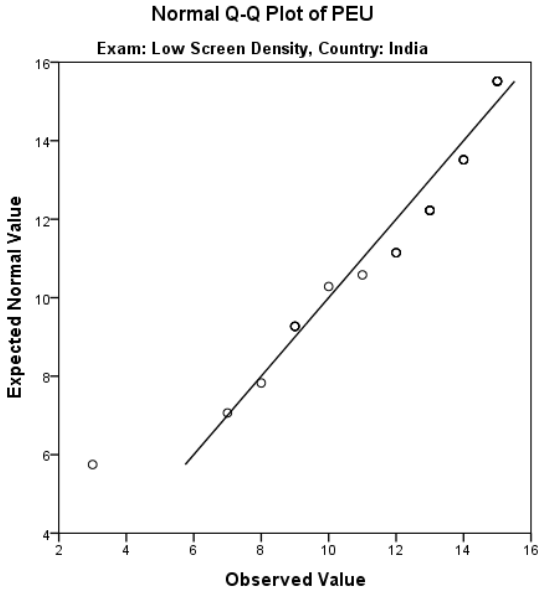


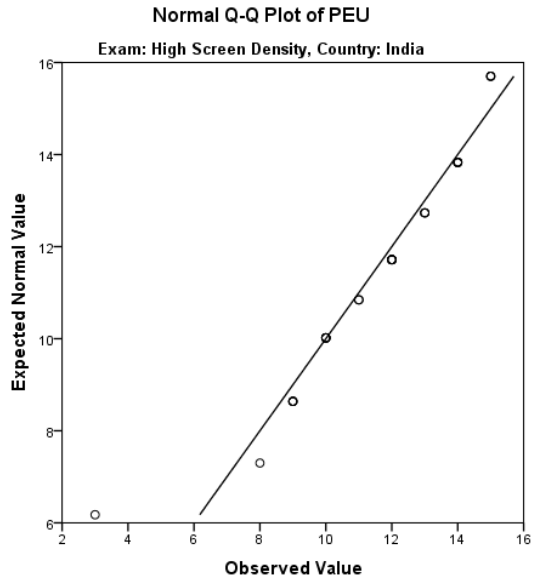


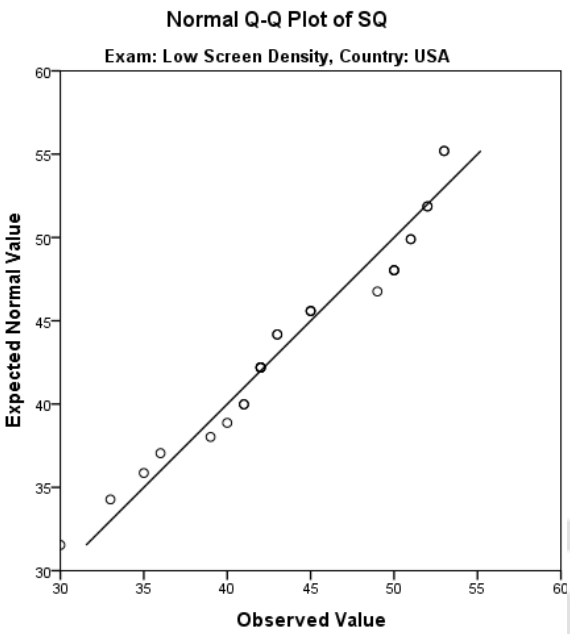


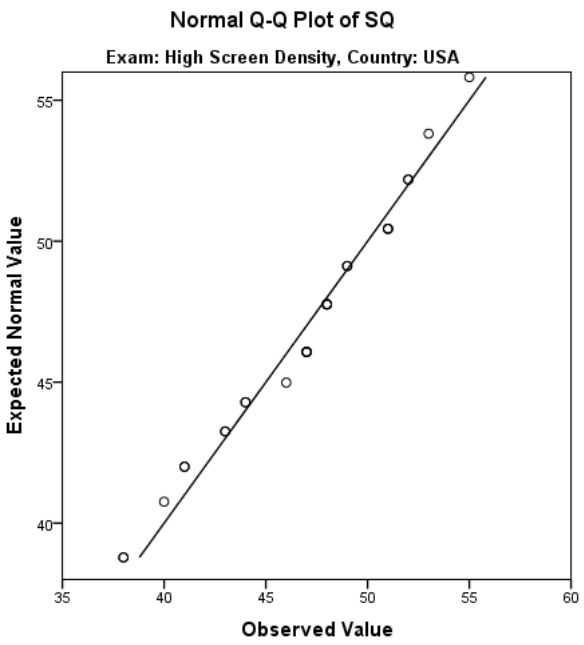


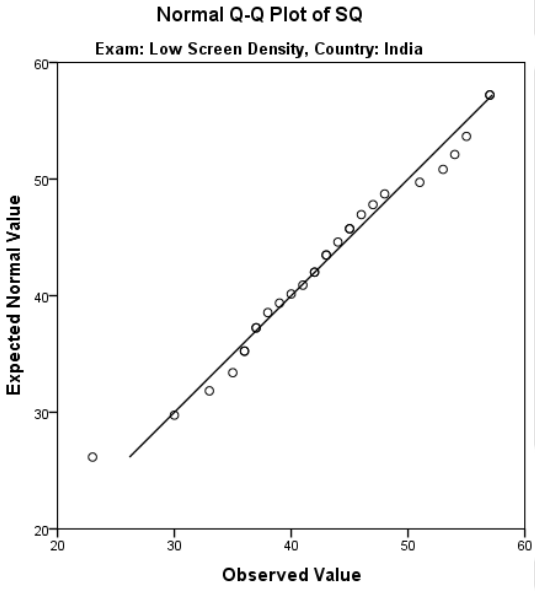


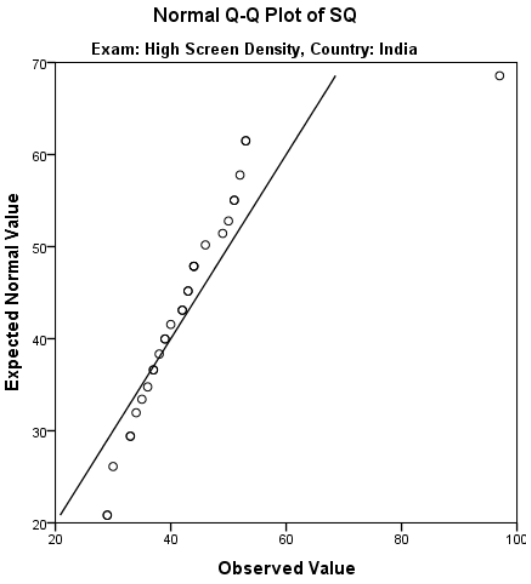


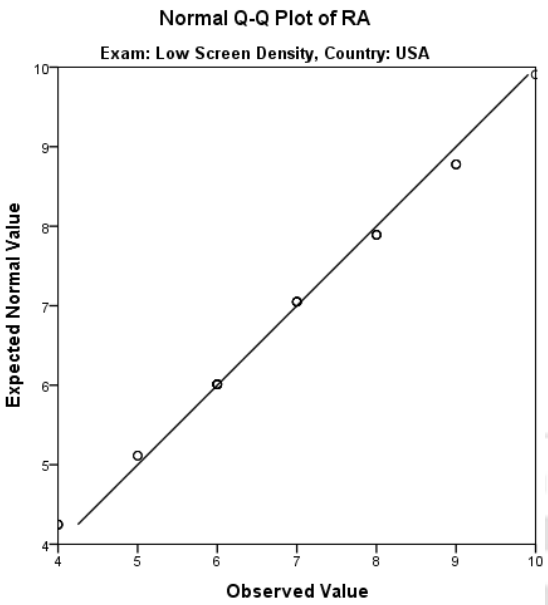


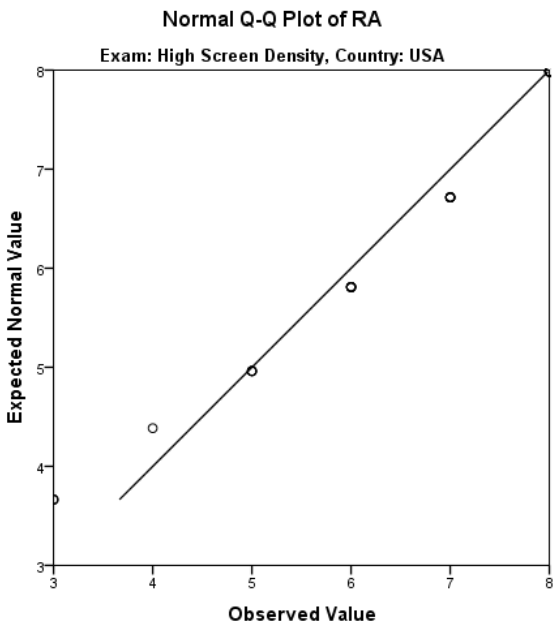


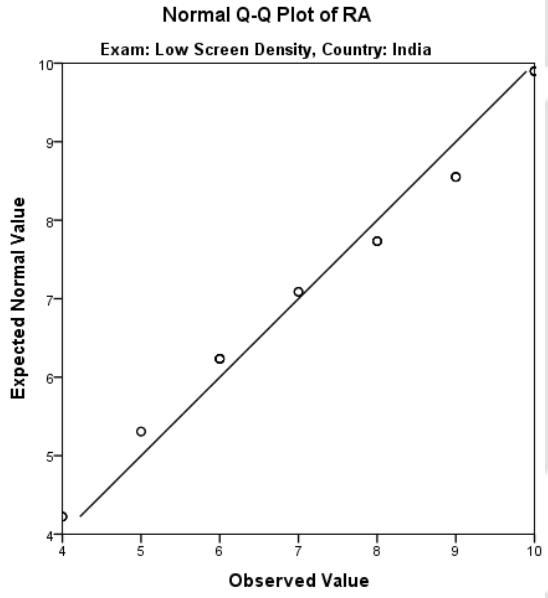


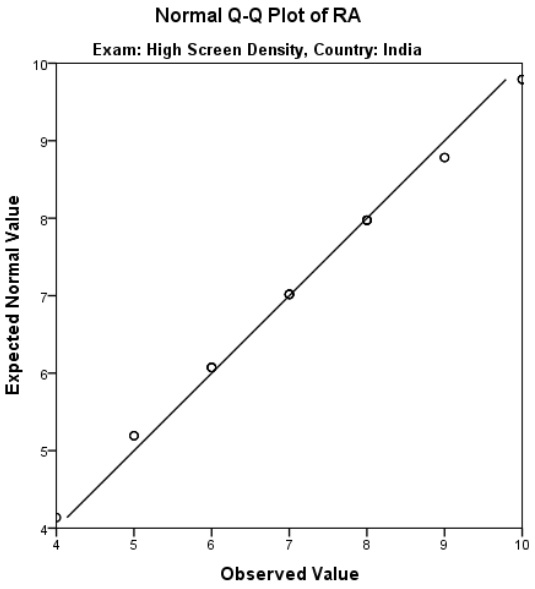


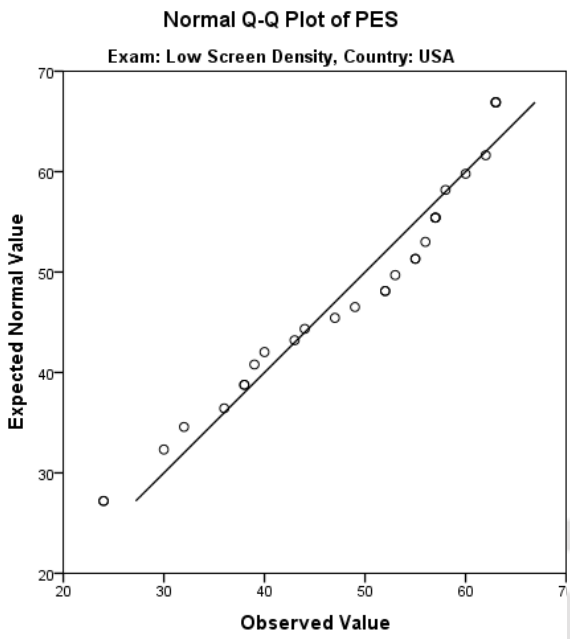


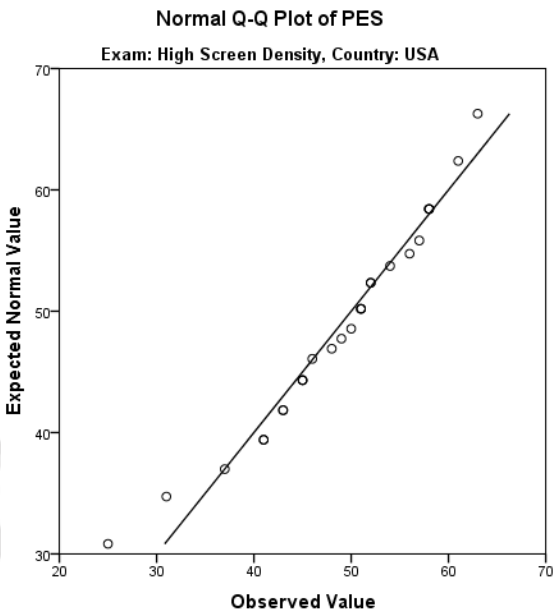


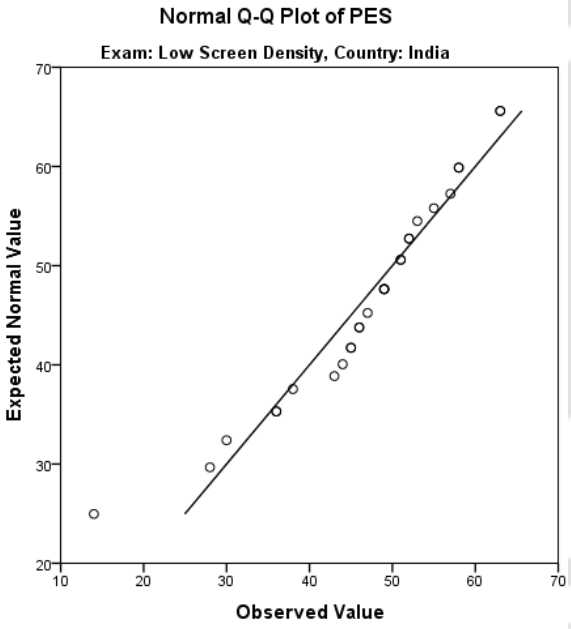


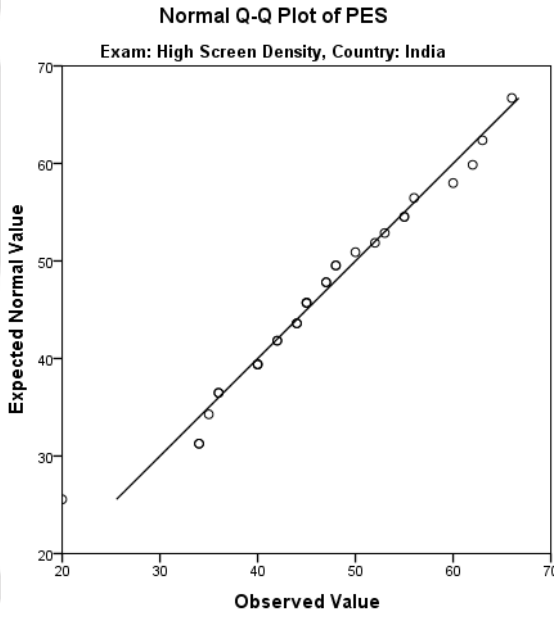


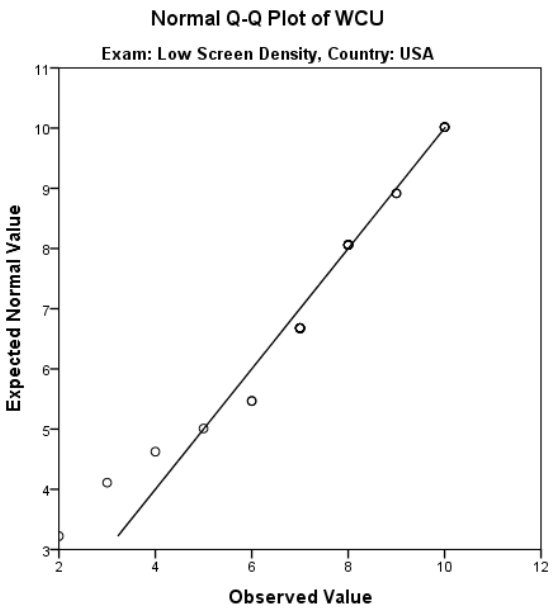


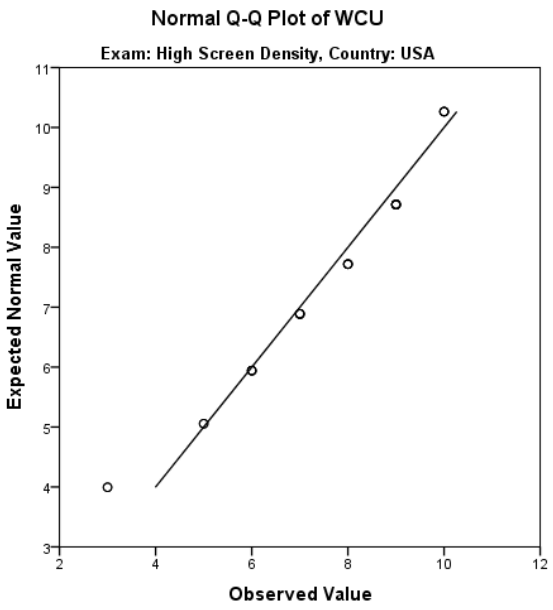


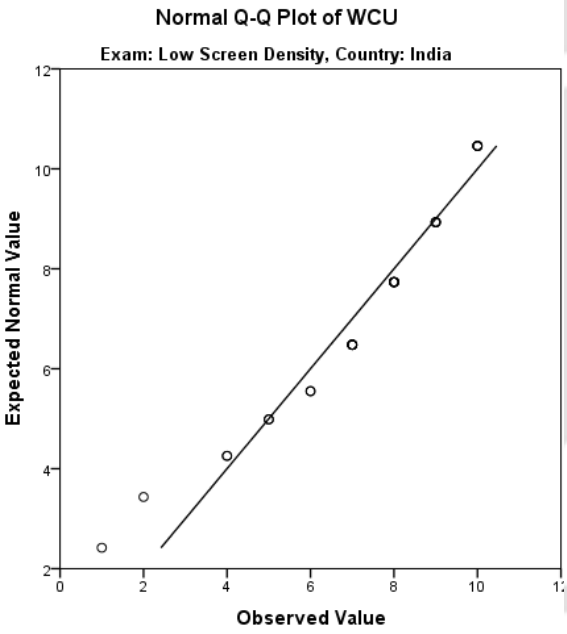


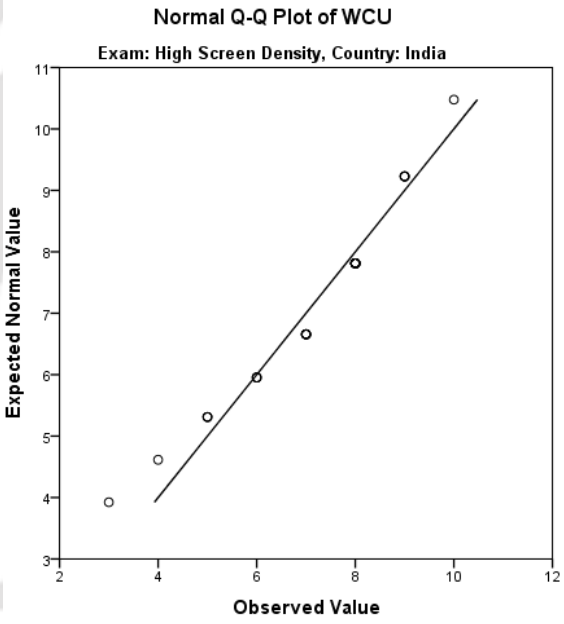


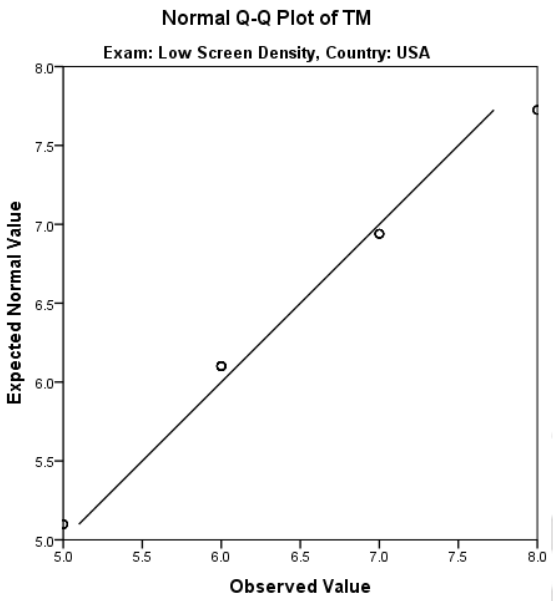


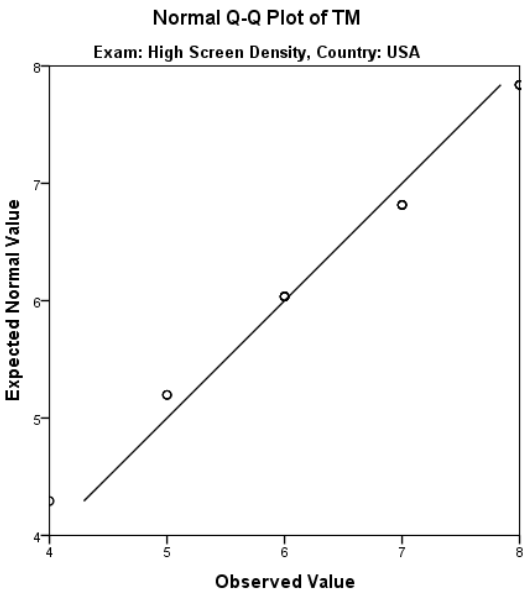


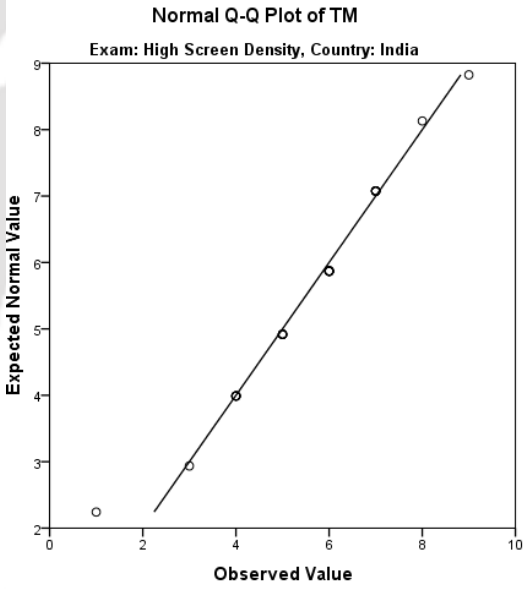


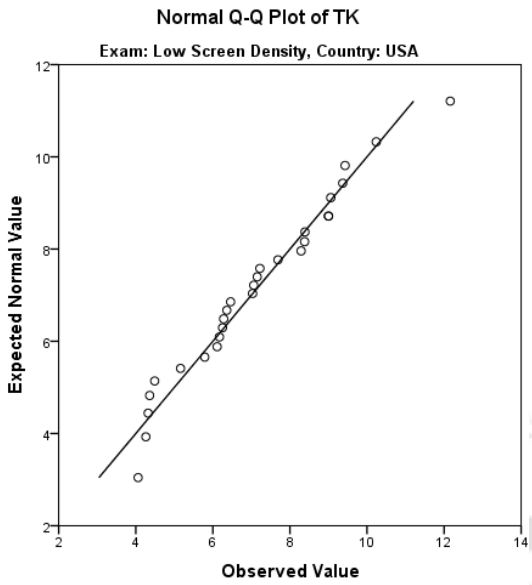


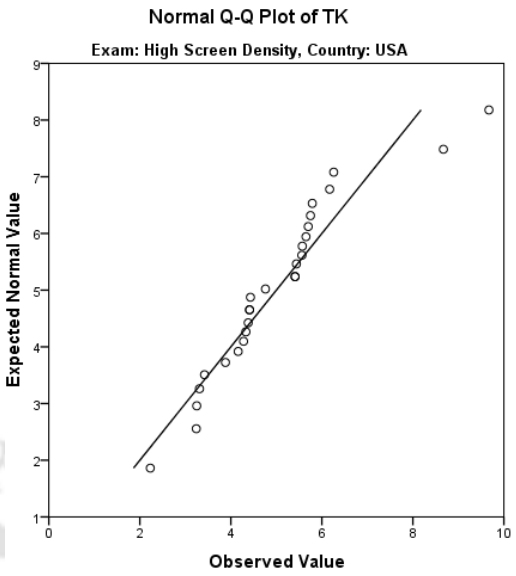


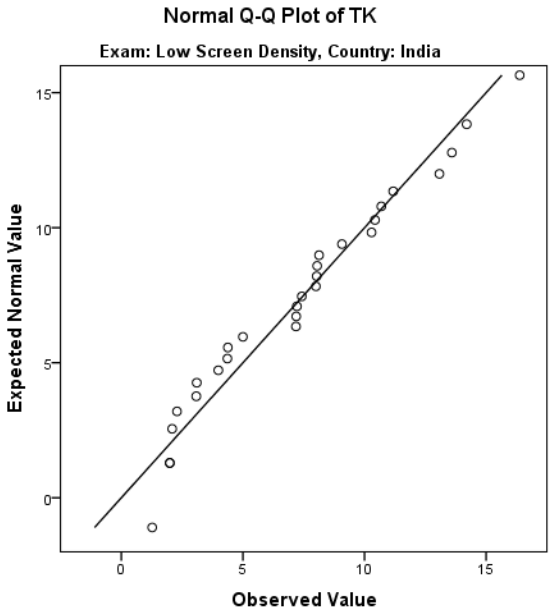












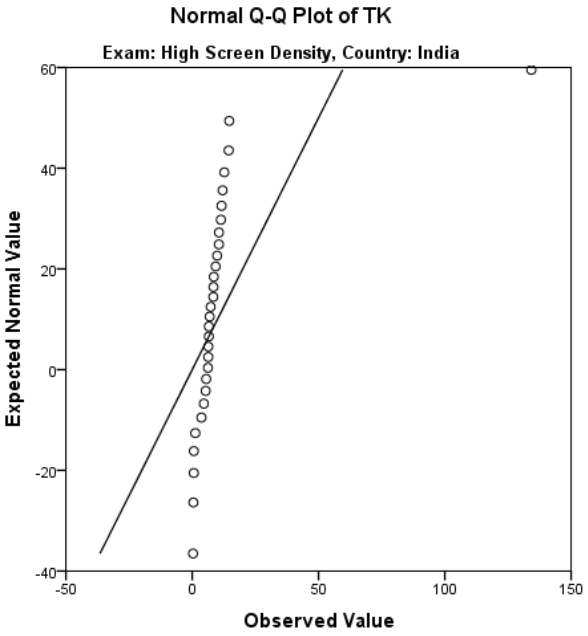


Table 3E.1

Box's test of equality of covariance matrices.

Box's M	559.39
F	2.82
df1	165
df2	25498.34
Sig.	.00

Box's test of equality of covariance matrices highlights that it is significant at $p<0.005$. Therefore, the assumption of homogeneity of variance-covariance of matrices has been violated. It suggests that the variance-covariance matrices among the dependent variables across the groups are not the same.

Table 3E.2

Levene's Test of Equality of Error Variances[a]

	F	**df1**	**df2**	**Sig.**
NASA TLX	1.53	3	109	.21
SUS	.75	3	109	.52
PEOU	.54	3	109	.65
PEU	1.59	3	109	.19
SQ	2.49	3	109	.06
RA	1.37	3	109	.25
PES	1.40	3	109	.24
WCU	.82	3	109	.48
TM	1.99	3	109	.11
TK	2.71	3	109	.04

Levene's test of equality of variances highlights that it is significant at $p<0.05$ for TK and not significant for the rest of the dependent measures. Therefore, the assumption of homogeneity of variances assumption has been violated for TK.

Table 3E.3

Univariate statistics of the dependent measures based on the independent factor – Exam and the interaction of the factors.

Source	Dependent Variable	Type III Sum of Squares	*df*	Mean Square	*F*	Sig.	Partial Eta Squared	Observed Power[k]
Exam	NASA TLX	354.57	1	354.57	1.44	.23	.01	.22
	SUS	221.20	1	221.20	1.33	.25	.01	.21
	PEOU	.40	1	.40	.05	.82	.00	.06
	PEU	.01	1	.01	.00	.98	.00	.05
	SQ	58.30	1	58.30	.80	.37	.01	.14
	RA	.95	1	.95	.30	.58	.00	.08
	PES	.07	1	.07	.00	.98	.00	.05
	WCU	.33	1	.33	.08	.77	.00	.06
	TM	.11	1	.11	.07	.80	.00	.06
	TK	31.76	1	31.76	.21	.65	.00	.07
Exam * Country	NASA TLX	.00	1	.00	.00	1.00	.00	.05
	SUS	45.53	1	45.53	.27	.60	.00	.08
	PEOU	19.47	1	19.47	2.61	.11	.02	.36
	PEU	3.18	1	3.18	.47	.49	.00	.10
	SQ	33.37	1	33.37	.46	.50	.00	.10
	RA	1.60	1	1.60	.51	.48	.00	.11
	PES	6.96	1	6.96	.06	.80	.00	.06
	WCU	.01	1	.01	.00	.97	.00	.05
	TM	1.42	1	1.42	.83	.36	.01	.15
	TK	283.04	1	283.04	1.85	.18	.02	.27

Table 3E.4

Pairwise comparisons of the dependent measures across India and USA

Dependent Variable	(I) Country	(J) Country	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
						Lower Bound	Upper Bound
SUS	USA	India	13.37*	2.43	.00	8.56	18.19
	India	USA	-13.37*	2.43	.00	-18.19	-8.56
RA	USA	India	-.81*	.33	.02	-1.47	-.15
	India	USA	.81*	.33	.02	.15	1.47
TM	USA	India	.76*	.25	.00	.27	1.25
	India	USA	-.76*	.25	.00	-1.25	-.27

Based on estimated marginal means

*The mean difference is significant at the .05 level.

[b]Adjustment for multiple comparisons: Bonferroni.

Table 3E.5

Correlational matrix of the dependent measures

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1. NASA TLX	1.00									
2. SUS	-.17	1.00								
3. PEOU	-.30**	.56**	1.00							
4. PEU	-.35**	.54**	.72**	1.00						
5. SQ	-.17	.44**	.49**	.56**	1.00					
6. RA	-.10	.18	.37**	.40**	.24*	1.00				
7. PES	-.32**	.49**	.54**	.59**	.44**	.38**	1.00			
8. WCU	-.18	.33**	.27**	.28**	.29**	.24**	.59**	1.00		
9. TM	-.14	.18	.15	.15	.17	-.12	.10	-.04	1.00	
10. TK	.04	-.07	-.08	-.09	-.05	-.02	-.02	-.05	.09	1.00

**p<0.01 (2-tailed), *p<0.05 (2-tailed).

Appendix 4

1. Disadvantages of CBAs

CBA systems have also been known to have a number of drawbacks because of which its large scale implementation has been hindered. Some of the limitations as highlighted by Funke in 1998 are listed below.

i. PC-based simulations are often so complex that even the developer of the system does not know what the correct or best solution will be for a given problem constellation.

ii. PC-based simulations produce a lot of behavioural data for most of which the psychological interpretation may be unclear.

iii. PC-based simulations cannot easily be evaluated with respect to their simulated domain validity.

iv. Results from PC-based simulations cannot easily be compared from one subject to another because the dynamic situations differ between subjects due to their different interventions.

v. PC-based simulations are low on the social dimension.

2. Theoretical implications of the Culture Variable

Literatures in Cultural studies present numerous perspectives through which culture and its dimensions have been defined (Straub et al. 2002). Schein in 1985 argued that basic assumptions are at the core of culture and represent the belief systems that individuals have toward human behaviour, relationships, reality and truth. These basic assumptions represent cognitive structures or interpretive schemes that people use to perceive situations and to make sense of ongoing events, activities, and human relationships, thereby forming the basis for collective action (Reichers and Schneider 1990; Sackmann 1992; Sapienza 1985; Van Maanen and Barley 1985). At the second level, values which represent manifestation of culture signify espoused beliefs identifying what is important to a particular cultural group. These values answer the question as to why people behave the way they do (Schein 1985). At the third level, artifacts and creations represent manifested culture that are more visible and observed easily. These artifacts may include things such as art, technology and visible and audible behaviour patterns as well as myths, heroes, language, rituals, and ceremony

(Pettigrew 1979). However, they are the hardest to decipher in terms of their underlying cultural meanings.

Modern day information technologies which can also be categorised as artifacts are not culturally neutral and may come to symbolize a host of different values driven by underlying assumptions and their meaning, use, and consequences (Coombs et al. 1992; Feld man and March 1981; Robey and Markus 1984; Scholz 1990). Among these three levels values are more easily studied than basic assumptions, which are invisible and preconscious and therefore not easily studied, as well as cultural artifacts (technology, art, visible and audible behaviors) that, being most visible, are not easily decipherable. Cultural values have a tight linkage with the subsequent behaviours and actions of social groups (Posner and Munson 1979). In this sense, values can be seen as a set of social norms that define the rules or context for social interaction through which people act and communicate (Delong and Fahey 2000; Keesing 1974; Nadler and Tushman 1988). These social norms have an impact on subsequent behaviours of firm members through acting as a means of social control that sets the expectations and boundaries of appropriate behaviours for members (O'Reilly and Chatman 1996). In this research investigation, we therefore take a position from the perspective of Cultural levels as highlighted by Schein in 1985. We argue that two students groups defined by their Culture markers would probably display different subsequent behaviours while interacting with different presentation formats in a Computer Based Assessment environment.

3. Importance of the research findings from the perspective of cultural variable

The results of the study highlights that students from different cultural groups interpret the functional features and the presentation format of the CBA system differentially. While Indians preferred single item presentation format US students preferred multiple item presentation formats. Likewise, in all the experiments that were carried out except the experiment with navigational mechanisms no two experiments recorded the same behavioural interaction of the two culturally diverse student groups. The findings of this cross-cultural study is important because it highlights that test performance of students across two diverse cultural groups can get affected because of the different social norms that are attached with these diverse groups and the way they interpret the presentation format.

GUI designers, while design interfaces for computer based assessment systems, should not provide uniform presentation formats for all student groups across different cultures. A uniform presentation format may hinder with the cognitive processes of the diverse student groups and affect their test performance while working with the system. It is therefore advised that GUI designers should provide flexible design solutions that can be customized according to students' cultural affiliations.

4. Cognitive Workload / Mental Load

Working memory is limited and deals with all conscious activities and the long term memory. Interactions between elements held in working memory require working memory capacity. For a particular intellectual skill, if working memory capacity is not freed for allowing the processes to occur, it would reduce the number of elements that can be dealt with simultaneously leading to over burdening of the working memory. This over burdening of the working memory is known as cognitive workload (Sweller, 1988).

5. Test of Homogeneity of Variance (Age and Marks obtained in last qualifying exam)

Experiment No.	Age		Marks Obtained in the Last Qualifying exam	
	Between Interface Feature	**Between Culture**	**Between Interface Feature**	**Between Culture**
1.	Levene's Test of Homogeneity of Variance is statistically significant at $p<0.005$	Levene's Test of Homogeneity of Variance is statistically significant at $p<0.005$	Levene's Test of Homogeneity of Variance is statistically not significant ($p>0.05$)	Levene's Test of Homogeneity of Variance is statistically not significant ($p>0.05$)
2.	Levene's Test of Homogeneity of Variance is statistically not significant ($p>0.05$)	Levene's Test of Homogeneity of Variance is statistically significant at $p<0.005$	Levene's Test of Homogeneity of Variance is statistically not significant ($p>0.05$)	Levene's Test of Homogeneity of Variance is statistically not significant ($p>0.05$)
3.	Levene's Test of Homogeneity of Variance is statistically not significant ($p>0.05$)	Levene's Test of Homogeneity of Variance is statistically significant at $p<0.005$	Levene's Test of Homogeneity of Variance is statistically not significant ($p>0.05$)	Levene's Test of Homogeneity of Variance is statistically not significant ($p>0.05$)
4.	Levene's Test of Homogeneity of Variance is statistically significant at $p<0.05$	Levene's Test of Homogeneity of Variance is statistically significant at $p<0.005$	Levene's Test of Homogeneity of Variance is statistically not significant ($p>0.05$)	Levene's Test of Homogeneity of Variance is statistically not significant ($p>0.05$)

5.	Levene's Test of Homogeneity of Variance is statistically significant at $p<0.05$	Levene's Test of Homogeneity of Variance is statistically significant at $p<0.005$	Levene's Test of Homogeneity of Variance is statistically not significant ($p>0.05$)	Levene's Test of Homogeneity of Variance is statistically not significant ($p>0.05$)

6. Experts vs Novice users

Experts have much experience and make fewer errors. From the perspective of Cognitive flexibility theory it can be argued that experts would outperform novices even in the case in which the hypertext provides a poorly defined structure of its content, because experts would adapt its flexible representation to match the organization of the system (Salmerón et al., 2005).

7. Future Scope

Given the findings of the research study it can be stated that future Computer based assessment systems should provide personalized/ customized features to the examinees.

CBA systems loaded with these cross cultural design features should be evaluated to access students' performance and their experience with the CBA systems in terms of usability/acceptability matrices.

8. Mental Model

Mental models are representations of reality that people use to understand specific phenomena. Johnson-Laird (1983) proposes mental models as the basic structure of cognition: "It is now plausible to suppose that mental models play a central and unifying role in representing objects, states of affairs".

List of Publications based on this thesis

1. Dhar, D., Adhikary, S., & Yammiyavar, P. (2012).An evaluation of the effect of navigational tools on cognitive load in a computer based test format. *Intelligent Human Computer Interaction (IHCI), 2012 4th International Conference on*, vol., no., pp.1, 6, 27-29 Dec. 2012, IEEE digital library, doi: 0.1109/IHCI.2012.6481875.

2. Dhar, D., & Yammiyavar, P. (2012).Design Approach for E-learning Systems: Should it be User Centered or Learner Centered. *Technology for Education (T4E), 2012 IEEE Fourth International Conference on* , vol., no., pp.239,240, 18-20 July 2012, doi: 10.1109/T4E.2012.57.

3. Dhar, D., & Yammiyavar, P. (2012).Influence of Culture on the design process of online learning environments. *Engineering Education: Innovative Practices and Future Trends (AICERA), 2012 IEEE International Conference on*, vol., no., pp.1, 6, 19-21 July 2012, doi: 10.1109/AICERA.2012.6306709.

4. Dhar, D., & Yammiyavar, P. (2012). Measuring Success of E-learning Applications: A Review. *Published in the proceedings of International Conference on Electrical Engineering and Computer Science, ICEECS 2012*.

5. Dhar, D., & Yammiyavar, P. (2012). Evaluating e-learning systems: Can we rely only on usability evaluation techniques. *Published in the proceedings of Symposium Human Computer Interaction in Virtual learning environments 2012, IITG*.

6. Yammiyavar, P., & Dhar, D. & (2011).Conceptualizing GUIs—A Case Study of an Online Aptitude Testing System. *Research into Design: Supporting Sustainable Product Development (ICORD), 2011 International Conference on*, Bangalore, India, 12.01.2011, ISBN: 978-981-08-7721-7.

Communicated to Journals

7. Dhar, D & Yammiyavar, P. ―Does scrolling affect students in computer based test format – A cross-cultural empirical investigation. *To Computers and education, Elsevier.*

8. Dhar, D & Yammiyavar, P. ―Effect of screen density in a computer based test format *To British journal of educational technology.*

9. Dhar, D & Yammiyavar, P. ―Question customization versus time customization-which format do students really prefer? *To Educational Technology Research and Development, Springer.*

Curriculum Vitae

Debayan Dhar

Born on 5th October, 1984, in Agartala, Tripura, India.

Education

1993 – 2001	Secondary Schooling, Ramakrishna Mission Vidyalaya, Vivek Nagar, Tripura.
2001 – 2007	Bachelor of Technology, Agricultural Engineering, North Eastern Regional Institute of Science and Technology, Nirjuli, Arunachal Pradesh.
2007 – 2009	Master of Design, Indian Institute of Technology Guwahati, Assam, India.
2010 – 2013	Research Scholar, PhD program, Department of Design, IIT Guwahati.
2012- 2013	Fulbright Fellow, University of Texas at Austin.

Jobs / Internships

June, 2009 – Sept., 2009	User Experience Analyst, Estuary Labs, Mumbai.
May, 2008 – June, 2008	Intern, Impelsys, Bangalore.

Research Interests

Human Computer Interaction

Usability Engineering

Instructional Design

Social Psychology

Awards

Fulbright – Nehru Doctoral and Professional Research Fellowship 2012- 2013, at the Department of Social Psychology, University of Texas at Austin, USA.