Identifying an effective framework for usability evaluation of Computer Supported Collaborative Learning System in educational settings

Eileen Huang

A Thesis
Submitted to
Auckland University of Technology
in partial fulfilment of the requirement for the degree of
Master of Computer and Information Sciences
In the School of Computing and Mathematical Sciences
Auckland University of Technology

2010

Primary Supervisor: Dave Parry

Authorship

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person (except where explicitly defined in the acknowledgements), nor material which to a substantial extent has been submitted for the award of any other degree or diploma of a university or institution of higher learning.

_____

Signature

Acknowledgments

This thesis was written at the School of Computing and Mathematical Sciences at the Auckland University of Technology.

I would like to take this opportunity to thank my supervisors, Dr Dave Parry and Mali Senapathi. I must first thank Dr Parry, my primary supervisor, for getting Ethics application approval, providing the valuable insights and feedbacks about my research, and guiding me where I should look into in order to keep me in the right direction toward the completion of the research in the last two years. Without Dr Parry's supervision, completion of this research would be impossible. I would like to thank Mali Senapathi, my secondary supervisor, for helping with completion of the research proposal, spending her time and working with Dr Parry to check the progress of the study so it could be carried out smoothly. Mali also sent me her master thesis which gave me some helpful ideas when I was writing this thesis.

Special thanks go out to Dr Tony Clear, who is the lecturer of Collaborative Computing. Dr Clear offered me the important opportunities of attending the classes and introducing my study to the students so I was able to recruit the participants for the case study. Without Dr Clear's great help, the case study could not be completed on time.

I would like to express many thanks to Ewing Caldwell for setting up OJS system for user testing; to Gordon Grimsey for helping arrange a laptop and a meeting room for user testing, to Diana Kassabova, my ex-supervisor and the proof-reader, for guiding me to complete the research proposal and for her hard work in proofreading and correcting the errors in the last draft of the thesis, to Boris Bacic and Shoba Tegginmath for taking part in the pilot study and sharing their important thoughts about the testing and the framework for OJS usability evaluation. I would like to thank Russel Pears and Robert Wellington for teaching me "Case study method", to Krassie Petrova, the programmer leader and a receptionist who worked in the afternoons for providing helpful services during my study at AUT.

I also would like to thank all students (the participants) in the Collaborative Computing class in Semester 2 2009 for taking part in user testing and making the valuable contribution to the study.

Finally I would like to send my special thanks to my friend Kiriwai Tapuke for spending her valuable time during her one-month vacation in Auckland on checking some chapters and helping me write the thesis in a better way in English.

Abstract

A review of previous studies has highlighted a gap between Computer Supported Collaborative Learning (CSCL) system usability evaluation (UE) in educational collaborative teaching and learning settings and groupware UE in general settings. The purpose of the research was to investigate this area further to identify a framework for CSCL system UE in educational settings. The framework should consist of the criteria that reflect the key features of CSCL system usability. Such a framework will then be capable of identifying the advantages and the disadvantages of a CSCL system's usability and its usability problems or issues.

Having considered a number of existing UE frameworks, a framework for CSCL system UE in educational settings was developed. The framework consists of 24 criteria grouped into six dimensions: Effectiveness, Efficiency, Collaborativity, Error tolerance, Universal accessibility, and Satisfaction.

The methodology for this research was designed as a two-year Case study (completed part-time) with six stages. It included user testing, one-to-one interviews, and questionnaires as the UE methods. The Open Journal System (OJS), a free online academic journal publishing system, was chosen as a collaborative learning (CL) system to test the developed framework. In this study, OJS had been set up for the Collaborative Computing (CC) paper in the School of Computing and Mathematical Sciences (SCMS) at the Auckland University of Technology (AUT). 18 participants took part in the study, and all except two academic staff were recruited from the postgraduate students in the CC class. The two academic staff joined in the Pilot study and 16 students/participants were divided into two groups – a student testing group and a follow-up group.

A 60-minute testing session was completed by each participant in the Pilot study group and the student testing group. This included a pre-test questionnaire, an asynchronous collaborative task (a peer review on OJS), a post-test questionnaire, and a one to one interview. 10 testing sessions were conducted. The participants in the follow-up group only completed the two questionnaires. Data was collected through a pilot study and the two groups mentioned previously.

The study has found that the defined criteria in the developed framework are important to UE and this framework is able to identify advantages and disadvantages of OJS. Limitations and problems in the research were identified. Future research should ensure that a larger sample size is used and user types are diversified, and that the framework's criteria can be best tested on several CSCL systems which support synchronous teamwork. Further investigation could be focused on how to determinate the weight for each of the six dimensions so that the framework can be improved and developed into an adaptable and effective assessment tool suitable for evaluating the system usability of a range of CSCL systems in educational settings.

Table of Contents

List of Figures

List of Tables

List of abbreviations

AUT: Auckland University of Technology

CC: Collaborative Computing (a course offered by SCMS at AUT)

CL: Collaborative Learning

CMC: Computer Mediated Communication

CSCL: Computer Supported Collaborative Learning

CSCW: Computer Supported Cooperative Work

FU: Follow-up group

HCI: Human Computer Interaction

ICC: Intraclass correlation

IR: Internal Reliability

IRR: Internal-Rater Reliability

OCS: Open Conference Systems

ODCS: Observer Data Collection Sheet

OJS: Open Journal Systems

PKP: Public Knowledge Project

SCMS: The School of Computing Mathematical Sciences

ST group: Student testing group

TG: Testing group including ST group and pilot study

TS: Task Sheet

TSH approach: Three-step Holistic approach

UE: Usability Evaluation

UI: User Interface

Chapter 1

1. Introduction

1.1 Background of this research

In the mid 1990s, the Internet and the associated world wide web (WWW) technologies hailed a new era of online and distance learning initiatives that included collaborative learning (Greenberg, Fitzpatrick, Gutwin, & Kaplan, 2000; Kildare, Williams, & Hartnett, 2006; Ross, Ramage, & Rogers, 1995). Various online applications and systems have been developed and used in educational settings over the years. One kind of such systems developed specifically for use in education to support group work is known as groupware or Computer-Supported Collaborative Learning (CSCL). Brinck (2005) and Wolz et al (1997) defined the CSCL system as any type of software or system that had been designed specifically to aid communication in groups and to facilitate group work. Such systems rely on modern computer networks, and have tools to transmit, store, annotate, and/or present information that has been created by one or more of a group's members.

Previous related studies have defined that collaborative learning (CL) and teaching requires a CSCL system that supports the creation of a group, structures learning activities, facilitates group interactions (Graham & Misanchuk, 2004) and allows the management of three crucial CL elements - participation, interaction, and synthesis (Ingram & Hathorn, 2004). So, an effective CSCL system should be capable of providing an online virtual place and good usability that helps teachers and students easily access it and work on these teaching and learning activities.

A good example of the CSCL system is WebCT. It is a Web-based educational software environment for customised design, delivery, and enhancement of educational and training courses delivered on the Web (Wolz, et al., 1997). Blackboard is another example and is currently used at Auckland University of Technology (AUT). In Blackboard, lecturers can manage their course materials, keep track of students' progress, and communicate with the students while students can access the system online at their preferred schedules from different locations. Both CSCL systems are institute-wide macro commercial systems.

The School of Computing and Mathematical Sciences (SCMS) at AUT offers postgraduate students a paper called "Collaborative Computing" (CC). The lecturer of this course would like to set up a low cost micro online system within the school where students could work collaboratively on their assignments in groups. On such a system, students would be able to communicate with teammates, peer review group assignments, modify and complete their group assignments.

The Open Journal System (OJS) is open source software created by the Public Knowledge Project (PKP) and released under the GNU General Public License[1]**.** The first PKP software was launched in 2001 (PKP, 2010). It is a free online journal publishing system, and offers PDF searching, a complete help manual, multiple rounds of reviewing, automated reminders, reviewer ratings, and a host of other features. It enables a single editor to manage publishing and index peer-reviewed journals over the Internet, and also supports an international team of editors, with shared responsibilities for a journal's multiple sections (Case & John, 2007; Willinsky, 2005). OJS is an asynchronous text-based online system. It was selected and set up as a collaborative teaching and learning system for the CC class in 2008 and 2009 respectively.

There had been questions about whether OJS was in fact a suitable system for collaborative teaching and learning. An evaluation of the system's usability and determination on whether or not it fit in with its intended purpose was necessary. Consequently, a framework for OJS CL usability evaluation (UE) needed to be identified or developed.

1.2 Motivation

In educational settings, Wolz et al. (1997) suggested that the ideal software or system for organising a course that contained both collaborative and other kinds of educational elements would "*be based on a simplified virtual reality approach*". Identification of appropriate online applications or tools to support effective communication and collaborative learning has been recognised as one of the key research topics in CSCL research (Li, Lau, Shih, & Li, 2008).

[1] **Type of open-source license:** GNU General Public License 2+ (Mark Cyzyk & Choudhury, 28 Apr. 2008)

The concept of usability has been given different meanings in different contexts. The ISO 9241-11 standard 1998 defines usability as "*the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use*" (UsabilityNet, 2006). This definition has been widely used and adapted in many different ways. According to Pinelle and Gutwin's suggestion (2008), a CSCL system's usability can be defined as "*the extent to which a groupware system allows teamwork and the activities of collaboration to occur - effectively, efficiently, and satisfactorily - for a particular group and a particular group activity*".

A CSCL system's UE requires specific measures for group/team activities, and addresses not only Human Computer Interaction (HCI) factors such as the effectiveness of interfaces and the quality of usability and interaction but also the aspects of pedagogy and learning from educational domains (Ssemugabi & Villiers, 2007b). Any evaluation should also consider the sequential and longitudinal characteristics of long-term activities such as information sharing, scheduling, role taking, synchronization, and allocation of resources (Neale, Carroll, & Rosson, 2004). Therefore a framework for assessing CSCL system UE should consist of the criteria which represent the key features of CSCL system usability defined above and be capable of identifying the issues or problems of a CSCL system's usability.

There was a paucity of research on developing a framework that would cover the key features of collaborative learning system usability and would be suitable for evaluating the usability of a CSCL system for collaborative teamwork and learning in educational settings. The previously developed UE frameworks did not include comprehensive criteria for evaluating a collaborative learning system's usability, particularly for evaluating the usability aspects supporting collaborative learning activities such as asynchronous or synchronous communication, monitoring, collaboration, and user/team management. Further research was necessary to address this.

With this in mind, the researcher decided to take on the challenge and make a contribution towards identifying and developing an effective framework for CSCL system UE. In this research, OJS was selected and set up as a CSCL system, and its usability was tested and evaluated by using the developed framework.

1.3 The objectives of this research

The aim of this research is to identify an effective framework for CSCL system UE in educational settings.

The objectives are:

1) to identify the important criteria for evaluating CSCL system usability in a collaborative teaching and learning environment;
2) to develop an framework for CSCL system UE in a collaborative teaching and learning environment;
3) to present some recommendations for future study in developing an effective framework for CSCL system UE in educational settings.

The research questions in this research are:

1) Does the proposed framework consist of important criteria for CSCL system UE in a collaborative teaching and learning environment?
2) Is the proposed framework capable of evaluating CSCL system usability in practice?
3) What should be improved in terms of the future study in developing a framework for CSCL system UE in educational settings?

1.4 Structure of the thesis

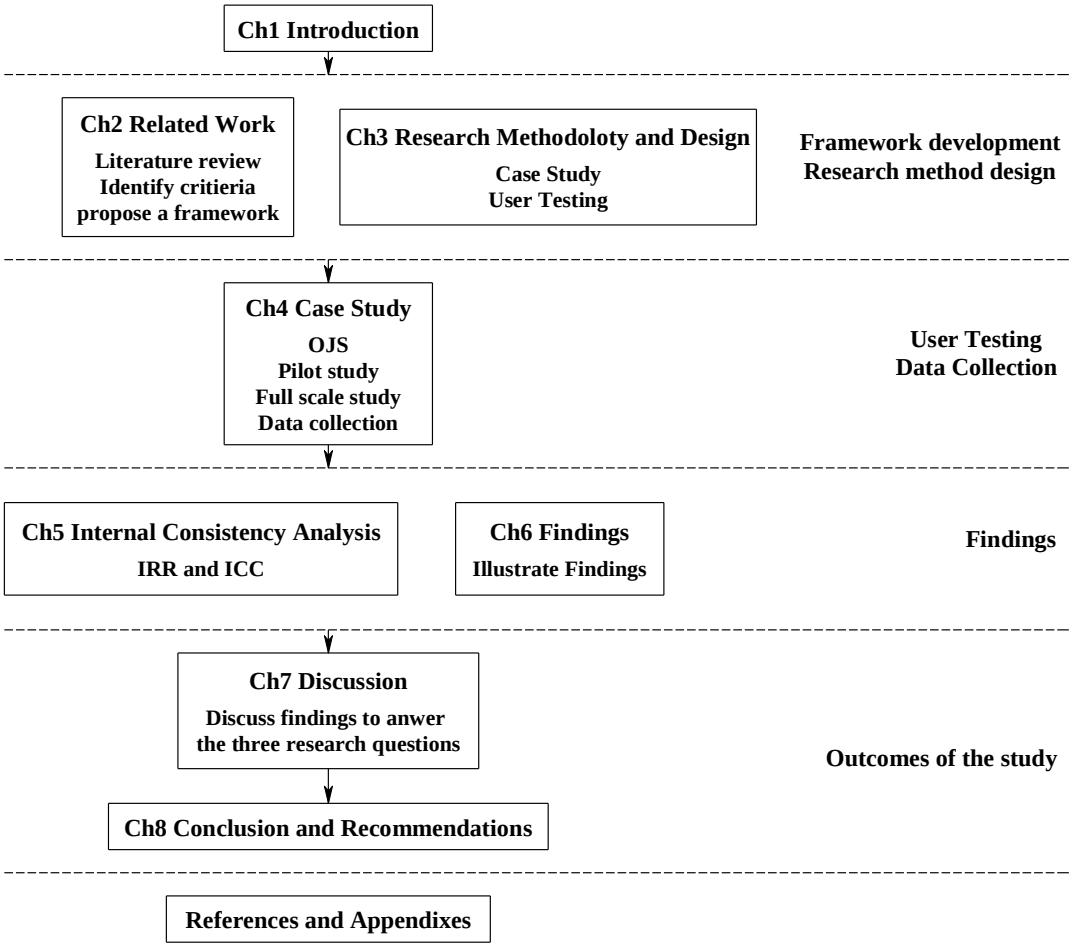


Figure 1-1 Structure of the thesis

This thesis consists of eight chapters and its structure is shown on Figure 1-1.
Chapter 1 introduces the background and motivation of the case study;

Chapter 2 provides a literature review of previous related studies on CL and CSCL, Usability and UE and CSCL system UE, UE international standards and UE frameworks. It then identifies the limitations and issues of previous frameworks and outlines the challenges and the opportunity of developing a framework for CSCL system UE. The chapter concludes with a proposed framework for CSCL system UE.

Chapter 3 presents the methodology and timeframe of the research. Case study was selected as the research method of this study. Three data collection methods were utilised – user testing, questionnaires and personal interviews.

In Chapter 4, a case study on OJS is presented in seven parts: OJS introduction, testing task design, trial testing, participant recruitment, pilot study, and full-scale study; the types of data collected from the study and the steps of the data process are described at the end of the chapter.

Chapter 5 introduces the theory about internal reliability and inter-rater reliability analysis. This chapter presents the findings from the reliability analysis of the six dimensions in the framework. A discussion of the findings and the main concerns from the analysis are also included.

Chapter 6 presents the findings drawn from the data that was collected from the Pre-test questionnaire and Post-test questionnaire, Task sheet and Observer data collection sheet, and the personal interviews. The main findings are summarised at the end of each subsection.

Chapter 7 discusses the findings and attempts to provide answers to the 3 research questions. It then reviews the methodology and the framework employed in the research and identifies the limitations, issues and problems relating to this research

Chapter 8 presents conclusions from the study and recommendations for future study.

A full list of references and appendices are provided at the end of this thesis.

Chapter 2

2. Related Work

This chapter looks into the related research on CL and CSCL system, usability and CSCL system usability, UE and CSCL system UE, and then illustrates some related UE frameworks. Finally a UE framework for CSCL system is proposed. Seven sections in this chapter are presented in an order shown as below.

2.1 Collaborative learning (CL)

2.2 Computer Supported Collaborative Learning (CSCL) System

2.3 Usability and CSCL system usability

2.4 UE and CSCL system UE

2.5 Review of related UE frameworks

2.6 The proposed framework for CSCL system UE

2.7 Summary

2.1 Collaborative learning (CL)

2.1.1 Definition of CL

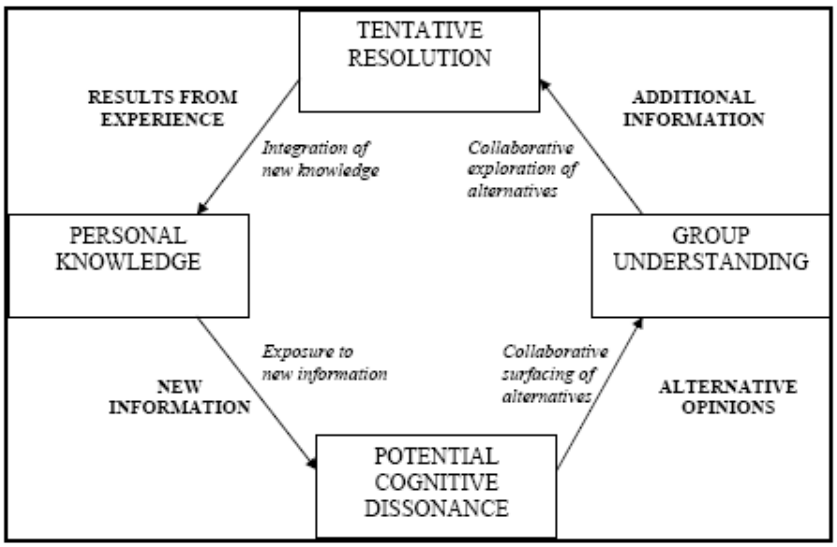
Collaborative Learning is "*a general term used for the description of educational practices based on the simultaneous cognitive and mental effort of multiple students or/and educators. Students share a common goal, depend on each other and are mutually responsible for their success or failure*" (Konstantinidis, Tsiatsos, & Pomportsis, 2009). It can be defined as the instructional use of small and medium-sized groups through which group members work together to maximise their own and each other's learning, to share understanding, and to achieve a common purpose or goal (Graham & Misanchuk, 2004; Wolz, et al., 1997) and is also "*viewed as a pedagogical method that can stimulate students to discuss information and problems from different perspectives, to elaborate and refine these in order to re-construct and co-construct (new) knowledge or to solve problems*" (Dimitracopoulou, 2005). Therefore, it is a social process that involves a cycle of individual and group knowledge construction (Wells & Brook, 2004). Group members (learners) are expected to exhibit considerable autonomy in their approach to a learning task (Koschmann 1994, as cited in Wells & Brook, 2004). They should be assigned to different roles in a learning task and have individual responsibility and accountability (Ewing & Miller, 2002).

So, in an educational setting, collaboration should bring learners together in a group (team) so that team members develop a better understanding of the topic at hand and are primarily concerned with the sharing, acquisition and creation of knowledge (Kildare, et al., 2006; Wolz, et al., 1997).

2.1.2 Previous studies on the CL process

Wells and Brook (2004) recommended a model, representing a summary of the various models of the phases of a CL cycle, as a framework for designing CL situations when applied to a tertiary learning environment. The model shown on Figure 2-1 defines that a CL cycle consists of five steps: existing personal knowledge, potential cognitive dissonance, group understanding, a tentative resolution, and then personal knowledge.

Figure 2-1 A Model of Collaborative Learning



(Wells & Brook, 2004)

Regarding the mechanics of collaboration, Steves, Morse, Gutwin, and Greenberg (2001) identified seven activities of teamwork collaboration in their study, i.e. *Explicit communication* (verbal, written, and gestural communication), *Implicit communication* (information from artifacts being manipulated, or information from others' movements and actions), *Coordination of action* (organising, avoiding conflicts and sharing resources and tools), *Planning, Monitoring* (& Tracking), *Assistance* (group members helping one another), and *Protection* (protecting group work). Pinelle and Gutwin (2002) advised that they are "*basic activities of shared work - the small-scale actions and interactions that group members must carry out in order to get a task done in a collaborative fashion"*.

Spada, Meier, Rummel, and Hauser (2005) combined a data-driven analysis of collaborative process with theoretical considerations and presented nine dimensions for

assessing the quality of various collaborative processes, that is, *"sustaining mutual understanding and coordinating communication refer to basic communication processes; information pooling and reaching consensus relevant for the construction and maintenance of a shared understanding; Task division, time management, and technical coordination reflecting the coordination of collaborative activities; shared task alignment and sustaining commitment refer to the motivational aspects"*.

The studies explained above suggested that the key CL processes should include: communication and team member interaction, task planning and monitoring and assistance, time management, team management, knowledge sharing and building, and content storage and protection.

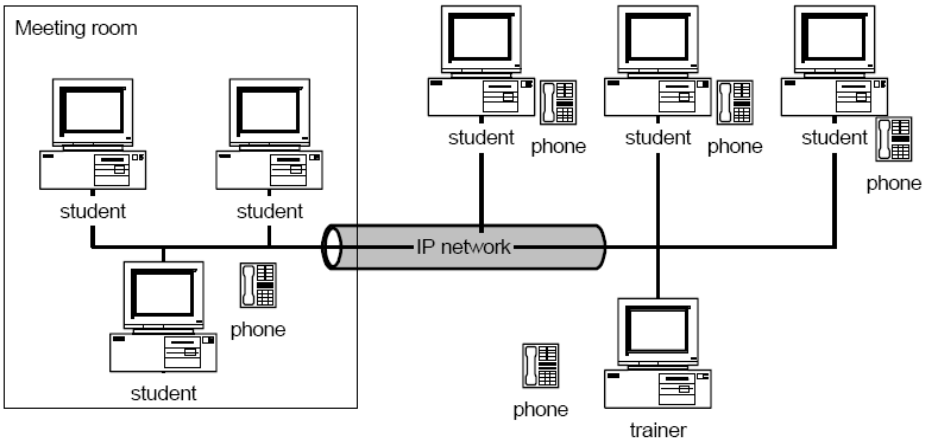
2.2 Computer Supported Collaborative Learning (CSCL) System

2.2.1 Definition of CSCL system

The Internet and WWW technologies started a new era of online learning/distance learning including collaborative learning in the mid-1990s (Kildare, et al., 2006; Li, et al., 2008). Various online applications and systems have been developed and used in educational settings since then. Kildare et al. (2006) commented that "*Online teams have become commonplace in our educational institutions and workplaces*". Groupware, Computer-Supported Cooperative Work (CSCW), Computer-mediated communication (CMC), and CSCL are the popular terms standing for the technologies that support online collaborative teaching and learning in educational settings.

CSCL system can be referred to as any types of software or systems that are designed for groups and for the communication and facilitating the work of groups, rely on modern computer networks, have some tools to transmit, store, annotate, or present information that has been created by one or more group members. The typical tools include email, awareness and notification systems, newsgroups, chat rooms, videoconferencing, and Internet-based audio application, and real time shared applications (such as collaborative writing or drawing) (Brinck, 2005; Wolz, et al., 1997). Figure 2-2 next page shows a computer network infrastructure for collaborative learning in different locations. Collaborative network and interface enable two or more people to work together concurrently on a task, even if they are separated by time and space. (Shneiderman & Plaisant, 2004).

Figure 2-2 A computer network infrastructure for collaborative learning



(Pfister, Wessner, Holmer, & Steinmetz, 1999)

One popular CSCL system example is WebCT, a Web-based educational software environment for customized design, delivery, and enhancement of educational and training courses delivered on the Web (Wolz, et al., 1997). Blackboard is another example. Instructors can manage their course materials, keep track of students' progress, and communicate with the students while students can access the systems at their preferred schedule through the Internet from different geographical locations. It has *"email and discussion forums to allow both students and instructors to interact with one another asynchronously*, (and also) *provides a text-based chatting function as well as a virtual classroom to facilitate synchronous collaborative learning*" (Kildare, et al., 2006).

There are two ways of classifying CSCL technologies. One way is based on the way of communication. CSCL systems can be categorised into two types:

1) *Asynchronous system*: allows participants to contribute to the discussion more equally. Full and free expression of ideas is possible (Ingram & Hathorn, 2004). Asynchronous communication tools maintain contact over time and share group documents (Bennett, 2004). "*In general, asynchronous learning is facilitated by emails and discussion forums."* (Kildare, et al., 2006)
2) *Synchronous system*: allows all participants to meet online at the same time, e.g. synchronous discussion suits brainstorming and quickly sharing ideas (Ingram & Hathorn, 2004). Nowadays, the synchronous communication tools like instant message, online chatting, video conference, etc, are widely used in higher educational settings.

The other way was suggested by Dimitracopoulou (2005). CSCL system can be divided into two categories according to the types of CL activities:

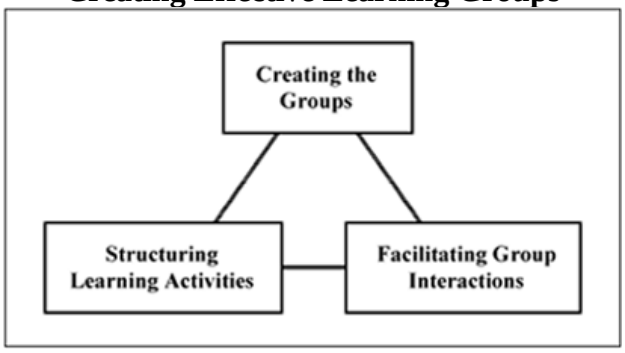
1) *Action-oriented collaborative systems*: based on the idea of starting from a student's actions, expressing and capturing the student's emerging knowledge and then making this knowledge-representation itself a subject of artifact-centered discourse. The underlying learning activities are synchronous;
2) *Text-production oriented systems*: invite students mainly to produce a written text or report in a collaborative or cooperative way. The underlying learning activities are asynchronous.

OJS used for this research is an asynchronous text-productions oriented system. More information about OJS will be introduced in Chapter 4.

2.2.2 Requirements of CSCL system

Wolz et al. (1997) suggested that a proper software or system for organising a course containing both collaborative and other kinds of educational elements will "*in the future be based on a simplified virtual reality approach*" and the educational culture should be taken into account - *"CMC tools need to be treated as an element (or elements) of a much larger learning culture. The focus needs to be on how these tools can be used to meet specific educational goals, not on the tools themselves*".

Figure 2-3 Three Important Elements in Creating Effective Learning Groups



(Graham & Misanchuk, 2004)

By looking into how to successfully facilitate group activities in a computer-mediated environment, Graham and Misanchuk (2004) identified three important areas that a CSCL system should support - "*1. Creating the groups, 2. Structuring the learning activities, and 3. Facilitating group interactions*" (see Figure 2-3), whereas Ingram and Hathorn (2004) defined three crucial elements of CL i.e. *"participation, interaction, and synthesis. Interaction requires that group members actively respond to one another. … The product that the group creates must represent a synthesis of ideas and input from all members*". Ardito, Costabile, Angeli, and Lanzilotti (2006) suggested that

"*As for any interactive system, the quality of the user interface is a primary requirement".*

Having considered the different views on CSCL above and the key CL processes identified in Section 2.1.2, this study suggests that a CSCL system should include:

1) *hosting* s*ystem* that would accommodate and manage a large number of users, including access control (Kildare, et al., 2006; Li, et al., 2008) ;
2) *content storage and delivery, and Course Management system* that supports different media delivery, and the management of the learning material (Dimitracopoulou, 2005; Kildare, et al., 2006; Li, et al., 2008; Wolz, et al., 1997);
3) *acknowledging Communication or Workspace awareness' functions* that would help group members know if other group members have read or received their communications (Dimitracopoulou, 2005; Graham & Misanchuk, 2004);
4) *various communication tools* that support asynchronous communication or synchronous communication or both, e.g. text production tools, dialogue tools for *discussion function*s (Dimitracopoulou, 2005; Ingram & Hathorn, 2004; Li, et al., 2008; Wolz, et al., 1997);
5) *analysis and metaanalysis tools* that support self-regulation and metacognition for students, and *teachers' tools* that supervise and analyse collaborative interactions either in an on-line or off-line mode (Dimitracopoulou, 2005);
6) *community level management system* that plans and monitors the activities and material produced amongst a wide community (Dimitracopoulou, 2005) and gives teacher control (Spada, et al., 2005; Wolz, et al., 1997);
7) *user account management system* that allows users to register and create account, to assign roles in a team, and to manage a team (Graham & Misanchuk, 2004) ;
8) *error prevention functions* that help users avoid making error via some tools such as error alerts, undo or reverse tools, etc (Shneiderman & Plaisant, 2004).
9) *security tools* that protect the contents stored in the system (Kildare, et al., 2006);
10) *help and advising functions* that lead to simple help systems or more advanced advising systems for students and teachers (Dimitracopoulou, 2005);
11) *the features of good user interface and ease of use* (Ardito, et al., 2006).

In summary, a CSCL system should have the capabilities of sharing and exchanging information, allowing group members participate, communicate and interact with one

another, producing the group's product(s) and achieving the group's common goal(s). Each requirement/capability determines its own aspect(s) of system usability.

2.3 Usability and CSCL system usability

2.3.1 Definition of Usability

The concept of usability has been given different meanings and standards in different contexts (González, Collazos, & Granollers, 2006; Green & Pearson, 2006; Mack & Nielsen, 1994). ISO 9241-11 standard 1998 defines usability as "*the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use*" (Green & Pearson, 2006; Ivory & Hearst, 2001; Scholtz, 2004; UsabilityNet, 2006). ISO/IEC 9126-1 standard in the product-oriented view defined it as "*the capability of the software product to be understood, learned, used and attractive to the user, when used under specified conditions*" (Bevan, 1995; UsabilityNet, 2006), whereas IEEE 1990 described usability as "*the ease with which a user can learn to operate, prepare inputs for, and interpret outputs of a system or component*" (Borgman & Rasmussen, 2005). Different researchers gave usability various meanings in their studies. For example, Rubin and Chisnell (2008) defined usability as "*a product or service should be useful, efficient, effective, satisfying, learnable and accessible*". About online application usability, Abascal, Arrue, and Vigo (2007) referred it as having "*a bearing on the users' effort in recognizing the logical concept and its applicability, learning its application, operation, and operation control. In addition, usability also refers to the facility with which the user can perform specific tasks in the web application*".

Currently, the widely accepted definition in a context of enterprise or other work-related applications is the one defined by ISO 9241-11 (Green & Pearson, 2006). Its three attributes (effectiveness, efficiency and satisfaction) were adopted in many definitions of usability. But some ambiguities have caused some usability problems by the definition itself in practice (Green & Pearson, 2006; Hu & Chang, 2006).

Brooke (1996) advised that it is impossible to specify the usability of a system without first defining who the intended users of the system are, what tasks the users will work on and complete, and what are the characteristics of the physical, organisational and social environment in which it will be used. If a system is used in a context which is an educational collaborative teaching and learning environment, then the definition of its

usability should not be exactly the same as the usability's definition for a stand-alone system used by a single user. In the next section, CSCL usability will be defined.

2.3.2 Definition of CSCL system usability

The features of CSCL system usability should be different from those of single user system usability. A CSCL system should allow users to share and exchange information, should support group members to participate, communicate and interact with one another, and produce the group's product(s) as discussed in sections 2.1.2 and 2.2.2. The CSCL usability should reflect the needs of team work and activities.

Pinelle and Gutwin (2008), and Pinelle, Gutwin, and Greenberg (2003) adapted ISO 9241-11 standard 1998 usability definition and defined CSCL or groupware usability as "*the extent to which a groupware system allows teamwork and the activities of collaboration to occur - effectively, efficiently, and satisfactorily - for a particular group and a particular group activity*". Brinck (2005) suggested seven attributes that should be considered: Group size, Ease-of-use, System responsiveness and reliability, Privacy and Security and Anonymity, Sharing Information, Identification and Accountability, and Control.

2.4 UE and CSCL system UE

In general, UE "*consists of methodologies for measuring the usability aspects of a system's user interface (UI) and identifying specific problems*" (Nielsen 1993, as cited in Ivory & Hearst, 2001). A CSCL system UE should consider the requirements of CSCL system defined on section 2.2.2, apply specific measures for team activities, and address not only Human Computer Interaction (HCI) factors such as the effectiveness of interfaces and the quality of usability and interaction but also the aspects of pedagogy and learning from educational domains (Ssemugabi & Villiers, 2007b). Evaluation should also consider the sequential and longitudinal characteristics of long-term activities such as information sharing, scheduling, role taking, synchronisation, and allocation of resources (Neale, et al., 2004).

2.5 Review of related UE frameworks

This section will review methodology-oriented frameworks, international standards and guideline, and some related UE frameworks, then identify the difficulties and issues

with CSCL system UE, and the need and the challenges for developing a framework for CSCL system UE.

2.5.1 Methodology-oriented frameworks

"Methodology-oriented frameworks describe the types of experiments and methodologies available to CSCW researchers. … are *useful for understanding the general types of evaluation possible, but provide little guidance for choosing among different types of methods*" (Neale, et al., 2004). This research has selected three most related frameworks and describes them below.

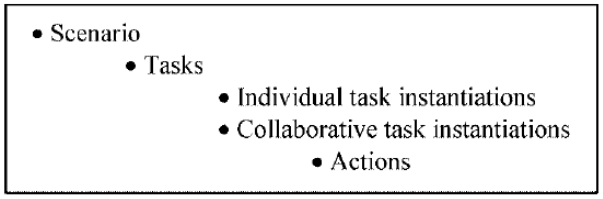
2.5.1.1 The Locales Framework

Greenberg, et al. (2000) defined that a locale is the actual site in which a group collaborates, people communicate, and groupwork is achieved. Locales should be dynamic so they can evolve along with the people, the artifacts, and the purposes that define them. The locales framework comprises five aspects: Locale foundations ("*a collection of people, resources and artifacts (tools, objects, information) in relation to the central purpose of the social world*"), Mutuality ("*those interactions within locales that maintain a sense of shared place*"), Individual view over multiple locales ("*acknowledges that individuals can be participating in many locales*"), Interaction trajectories *("how courses of action evolve over time*"), and Civic structures ("*how interactions fit within a broader communal level*"). Providing awareness within locales is very important as it helps people maintain a sense of shared place and that keeps them informed about shared activity. Mutuality includes one person's awareness of others, where things are located, and how things are changing (Gutwin & Greenberg, 1999).

2.5.1.2 Collaboration Usability Analysis. (CUA)

Pinelle, et al. (2003) argued that the existing task analysis schemes were not well suitable to the needs of groupware evaluation as they either do not deal with collaboration issues, or adequately represent the variability inherent in group work. They developed a new modelling technique called "Collaboration Usability Analysis" (CUA) (see Figure 2-4) and considered it as a conceptual CSCW framework (Pinelle & Gutwin, 2002).

Figure 2-4 The component hierarchy in CUA task model



(Pinelle, et al., 2003)

The framework divides group activity into two areas – *taskwork* completed by each individual group member and *teamwork* completed by a group. It is based on a hierarchical task model that represents the procedural elements of a group task in a shared workspace. The collaborative activities can be analysed and specified by decomposing them into smaller, mechanical units. The mechanics cover two general types of activity: communication and coordination. Communication is broken into two categories: explicit communication and information gathering. Coordination is broken into two categories: shared access and transfer (Pinelle, et al., 2003; Steves, et al., 2001). The framework has offered a way of understanding a system, analysing the activities run on the system, and evaluating the system usability.

2.5.1.3 Metrics for usability

López-Jaquero, Montero, Fernández-Caballero, and Lozano (2003) pointed out that evaluation can fall into three broad categories: preference metrics, which "*quantify the subjective evaluations and preferences of users*", performance metrics, which "*measure the actual use of working software*", and predictive metrics, or design metrics, which "*assess the quality of designs and prototypes*". They recommended that researchers should focus on preference and performance metrics because one of the most popular ways to assess usability is to use preference metrics while performance metrics are "*especially useful for assessing overall usability*" and "*most of them can be evaluated at run time in a simple manner*".

2.5.1.4 Applications of the three frameworks

The three methodology-oriented frameworks explained above supply useful structures of designing and developing an effective and practical framework for CSCL system UE. For this research, the Locales Framework defines the five aspects of UE a groupware including CSCL system; CUA recommends a way of analysing individual and team tasks undertaken in OJS which was selected for usability testing in this study; Metrics for Usability offered some types of metrics which can be chosen as some usability measures and included in a framework for CSCL system UE.

2.5.2 International standards and guidelines

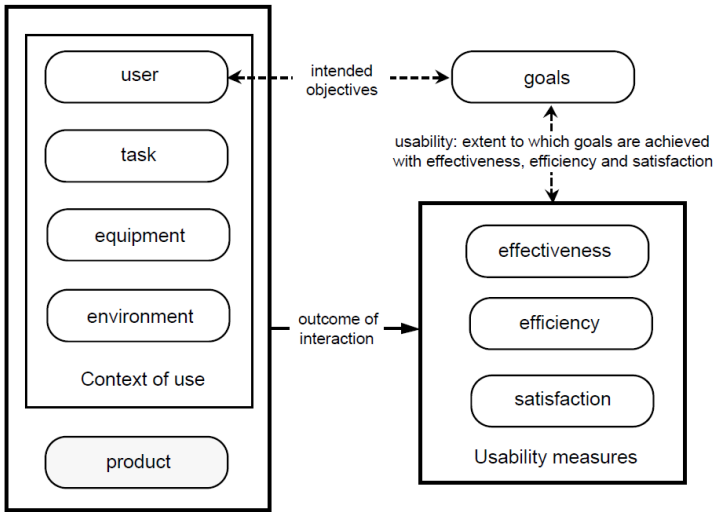
2.5.2.1 Three International standards for UE

Three international standards for UE are commonly used in system UE. The standards "*are intended to provide guidelines and general principles for planning and conducting evaluation*" (Marghescu, 2009). The first one is W3C web standards which include different guidelines for various aspects of online systems. For example, the current

version of Web Content Accessibility Guidelines (WCAG) is WCAG 2.0 and was published on 11 Dec 2008. It has 12 sub-guidelines that are organised under 4 principles: *Perceivable* (Information and user interface components)**,** *Operable* (User interface components and navigation)**,** *Understandable* (Information and the operation of user interface)**,** and *Robust* (Content) (Henry, 2008). This standard focuses more on user interface than the other two standards explained below.

The second standard is CISU-R standard, developed by U.S. National Institute of Standards and Technology (NIST) (version v0.90 was presented in March 2007). It specifies usability requirements and includes three types of information: the context of use, performance and satisfaction criteria, and the test method and context of testing. The measures of usability consist of performance measures including *Effectiveness* (Task completion rate, errors, assists) and *Efficiency* **(**relative user efficiency, completion rate), and *satisfaction* measures using an established questionnaire. The standard is consistent with the definition of usability in ISO 9241-11 and can be applied to hardware and software products (Henry, 2008; NIST, 2007).

Figure 2-5 ISO 9241-11 Usability Framework



(Bevan, 1995)

The third standard is ISO 9241-11. It is the most widely used standard in usability evaluation and its study. Usability is dependent on the context of use which consists of the users, tasks, equipment (hardware, software and materials), and the physical and organisational environments that may all influence the usability of a product (see Figure 2-5). The measures of usability include:

1) *effectiveness*: accuracy, completeness / goal achievement, and the quality of the output of the tasks;
2) *efficiency*: time, speed of completeness, the resources expended in achieving goals;

3) *satisfaction*: users' *subjective reactions* to using the system.

(Bevan, 1995; Brooke, 1996; Skov & Stage, 2005)

The precise measures to be used within different contexts can vary widely (Brooke, 1996). Efficiency, effectiveness, and satisfaction should be considered independent aspects of usability, unless domain specific studies suggest otherwise (Frøkjær, Hertzum, & Hornbæk, 2000).

2.5.2.2 Two guidelines for system design

Two guidelines provide the requirements for system design in details. The first one is "Eight Golden Rules of User Interface Design*"* and was recommended by Shneiderman and Plaisant (2004). The rules and principles were derived from experience and refined over two decades and had been well received as a useful guide to students and designers. The rules are *Strive for consistency (e.g.* colour, layout, capitalisation, fonts, etc), *Cater to universal usability* (e.g. considering diverse users and design for plasticity, adding different features for novices and experts), *Offer informative feedback* (e.g. showing changes explicitly), *Design dialogs to yield closure* (e.g. showing sequences of actions), *Prevent errors* (e.g. validation of input on a form), *Permit easy reversal of actions* (e.g. undo), *Support internal locus of control* (e.g. modification and configuration done by experienced operators), *Reduce short-term memory load* (e.g. displays kept simple, multiple-page displays be consolidated).

The other one is called as "Seven Universal Design Principles" and consists of: "*Equitable Use, Flexibility in Use, Simple and Intuitive, Perceptible Information, Tolerance for Error, Low Physical Effort, Size and Space for Approach and Use"*. The last principle is more related to hardware (CUD, 2008).

2.5.2.3 Applications of UE standards and the guidelines

The standards and guidelines described above have identified some essential and critical aspects of system usability, such as Effectiveness, Efficiency, and Satisfaction from the standards, Error Prevention and Tolerance, and Universal Usability from the guidelines. They could be adopted as dimensions or criteria for developing a framework for CSCL system UE.

However, none of them cover the aspects related to collaborative activities such as collaborative teaching and learning activities. In next section, the thesis will review

some related UE frameworks in order to identify the important criteria for evaluating CSCL system usability.

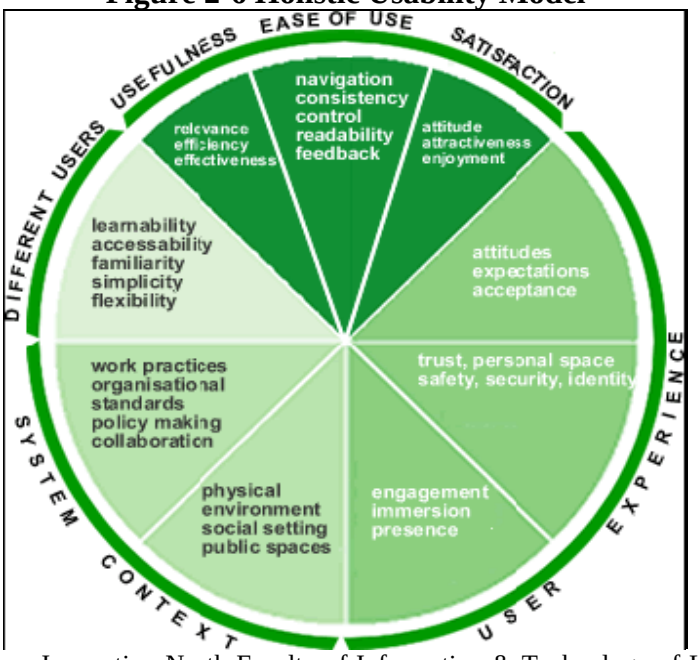
2.5.3 Related UE frameworks

Appendix 1 lists several practical UE frameworks produced in the last 10 years. Different frameworks have different attributes or criteria for evaluating different systems in different contexts. This research has carefully compared them and their contexts, and selected the two related frameworks described below as the main references for developing a framework for CSCL system UE.

2.5.3.1 Holistic Usability Framework

The refined framework was used to evaluate two existing distributed simulation systems (DSSs), measuring their holistic usability. It consists of six dimensions: end user needs, end user interface(s), programming, installation, training, and documentation. Each dimension has its own attributes. The total number of criteria of this framework is 55 (Dawson, 2006). (see Figure 2-6, Framework 13 in Appendix 1).

Figure 2-6 Holistic Usability Model



(**Source**: Innovation North Faculty of Information & Technology of Leeds Metropolitan University, (as cited in Dawson, 2006)

2.5.3.2 A multi-faceted framework - for web-based learning

This framework is "*an effort to integrate usability and learning, particularly in the context of a constructivist approach. They are appropriate for the evaluation of web-based e-learning*" (Ssemugabi & Villiers, 2007a) (more details are shown on Framework 18 in Appendix 1). It includes 20 attributes and three categories - *Category 1:* General interface usability criteria (based on Nielsen's heuristics, modified for e-

learning context); *Category 2*: Website-specific criteria for educational websites; and *Category 3:* Learner-centred instructional design, grounded in learning theory, aiming for effective learning.

2.5.3.3 Applications of the two related UE frameworks

The contexts of the two frameworks outlined above are similar to the context of this research. The frameworks consist of several aspects of system usability and include more than 20 criteria for UE. Several criteria (such as navigation, organisation, learner control, feedback, guidance, and assessment) were used to measure a system's usability supporting teamwork or online learning system's usability. However, the frameworks were not designed and developed specifically for CSCL system UE. Therefore, it is inappropriate to adopt either of the two frameworks or combine both to form a framework for CSCL system UE.

2.5.4 Issues and difficulties with CSCL system UE

Lack of standard assessment criteria for reliably comparing usability evaluation methods is an important issue in HCI knowledge (Chattratichart & Brodie, 2004). Although ISO standard has been widely used in UE, "*the ISO 9241-11 definition is not a broad enough view of human interaction to describe the usability goals of either the users, or the business"* (Green & Pearson, 2006). The methods by which effectiveness, efficiency, and user satisfaction were measured vary greatly, and are inconsistent and difficult to compare. "*There is a large body of practitioner literature, which was fragmented, spanning across many disciplines, with the approach varying depending on the viewpoint taken*" (Dawson, 2006). Usability evaluation results vary dramatically when applied by different evaluators (Zhang, 2007).

Undoubtedly, evaluating distributed CSCW and CSCL applications is a difficult endeavour (Greenberg, et al., 2000; Neale, et al., 2004; Ross, et al., 1995). Neale et al. (2004) identified three problems that made this type of evaluation difficult: "(1) *The logistics in carrying out distributed evaluation are difficult; (2) There are a greater number of variables to consider, and they are more complex; and (3) evaluation in much of CSCW needs to focus on validating the reengineering of group work based on CSCW concepts*. … *The variables of complex synchronous and asynchronous interaction have not been fully identified or understood"*. CSCW involves multiple humans interacting with networked systems. This makes the problem at least an order of magnitude more complex than single user systems (Damianos, et al., 2000). In addition,

"*a comprehensive evaluation of a web-based learning environment needs to consider both the technical as well as the pedagogical aspects of the system. This may be seen to add an extra level of complexity to evaluations of conventional educational systems*" (Hollingsed & Novick, 2007).

2.5.5 The need for developing a CSCL UE framework

Several UE frameworks have been developed and have various numbers of criteria (See Appendix 1). For example, three criteria were suggested by Ivory and Hearst (2001) while 55 criteria were identified in Dawson's study (2006). Three essential dimensions (effectiveness, efficiency, and satisfaction) defined by ISO 9241-11 were included in many frameworks. Other dimensions and their criteria were further defined based on the systems' contexts and features. According to the list of UE frameworks shown on Appendix 1, the most often used dimensions and criteria are satisfaction, effectiveness, efficiency, learnability, memorability, and error management/tolerance; the often used dimensions and criteria are ease of use, user control, consistency/standards, help and documentation, visibility, speed, and accessibility.

However, none of the related frameworks shown in Appendix 1 includes comprehensive criteria for evaluating CSCL system usability, particularly for evaluating the usability aspects (listed on section 2.2.2) supporting collaborative learning activities such as asynchronous or synchronous communication, planning and monitoring, user/team management, security, file storage and sharing, and so on. Sheard and Markham (2005) noted that "*a search of the literature has shown a scarcity of systematic evaluative studies of web-based learning environments*". Little has been done to critically examine web-based e-learning applications' usability although there has been widespread use of the applications for distance and classroom learning over the past decade and usability evaluation methods have a long history of research (Schmettow & Vietze, 2008; Ssemugabi & Villiers, 2007b). Much of the prerequisite research used to understand CSCW has not been based on teams (Neale, et al., 2004). The determinants of success in Collaborative interfaces are still not clear (Shneiderman & Plaisant, 2004).

Neale et al., (2004) suggested that researchers must consider the aggregate of the factors such as communication, coordination, cooperation, awareness when evaluating a CSCW system for teams. Neale and Carroll (1999) advised that when evaluating CSCL system, one must consider user interface usability issues, coordinated multi-user computer issues, learning efficacy in general, cooperative aspects of group learning specifically,

and the larger context of the classroom(s) in which the previous issues are situated. Therefore, an effective CSCL system UE framework applied in collaborative teaching and learning settings needs to be looked into and developed. New research in UE should aim to develop a UE framework that includes the criteria that would be suitable for evaluating CSCL system usability in educational settings.

2.5.6 The challenges of developing a framework for CSCL system UE

"*The development of valid and reliable outcome measures is a prerequisite for assessing overall system usability*" (McGee, 2004). As discussed in sections 2.5.1 - 2.5.5, many researchers have made their contributions to the methodologies of CSCL and developing frameworks for UE. However, the existing UE frameworks do not consist of comprehensive criteria for CSCL system UE. Effectively evaluating CSCL system usability is not easy. So, one challenge is to identify important criteria and then develop an effective framework that should be capable of identifying the advantages and disadvantage and problems of a CSCL system when it is applied to the CSCL system UE in an educational CL setting.

A UE framework includes a number of criteria and dimensions. One criterion or a dimension may be more important than others. Dawson (2006) suggested that the importance of each criterion could be affected by types of users (novice or expert), and the features and functions of a system. For example, for a game system, satisfaction may be deemed more important than that for a military battle simulation. On an online system, ease of use may be more important than other attributes of usability because often there is no opportunity to train users to use the system (Green & Pearson, 2006). Considering more of practical evaluation, efficiency and satisfaction should be focused on. The measures for these two attributes should include time to learn, speed of performance, rate of errors by users, retention over time, and subjective satisfaction (Shneiderman & Plaisant, 2004). Therefore, how to weigh various criteria or dimensions in a UE framework is another challenge.

2.5.7 Summary

In summary, an effective framework for CSCL system UE had not been developed yet. This research had taken the first challenge and the opportunity of this study, reviewed the related studies on UE frameworks, and identified important criteria to develop a framework for CSCL UE. The proposed framework will be illustrated in the next section.

2.6 The proposed framework for CSCL system UE

Following literature studies on the UE methodologies and frameworks, this research had considered the features of the collaborative learning in a CSCL system, the requirements of a CSCL system, and CSCL usability as well (explained in the sections 2.1, 2.2.2, 2.3.2 respectively) and developed a framework for CSCL system UE in educational settings. ISO 9241-11 standard was adopted as the scaffolding and some criteria were selected from the existing related frameworks discussed in Sections 2.5.1.4, 2.5.2.3 and 2.5.3.3.

Figure 2-7 The proposed framework for CSCL system UE

This framework consists of six dimensions and 24 criteria (see Figure 2-7). Each dimension has its own criteria which were carefully selected according to the review of previous related studies (shown in Table 2-1 on pg26 and Appendix 1) and the information presented in the sections above in this chapter. The six dimensions are:

1) *Effectiveness* means the completeness and accuracy with which users achieve their goals (Hu & Chang, 2006; Quesenbery, 2003). This dimension has four criteria:

Completeness, Visibility, Organisation and design, and Navigability. Good visibility and navigability and proper organisation and interface design are essential for users to be able to complete tasks and to achieve their team's goals when using a CSCL system.

2) *Efficiency* aims to measure the speed (with accuracy) with which users can complete their tasks. It is usually a measure of time (Rubin & Chisnell, 2008). This dimension consists of four criteria: Speed, Familiarity/Consistency/Standards, Flexibility/Adaptability/Configurability, and Effort. Consistency in design and actions of using a system and adopting a standard would help users understand and learn a system quickly. Flexibility and Adaptability and Configurability are different aspects of usability but aim to achieve the same goal – meeting users' requirements and providing users with a fast and easy access to a CSCL system. If a system is flexible, it should be capable of being configured and customised to what users want and what they are used to. Therefore, users would be able to complete tasks/jobs with little effort.
3) *Collaborativity* dimension aims to evaluate a system's usability aspects that support collaborative teaching and learning. It has eight criteria: User Management, Awareness, Communication, User Control/Moderator and Teacher Control, File/Content Sharing/Management, Process Tracking/Automated Notification, File/Content Protection, and Security.

   A CSCL system should have the utility to allow users to administrate and manage team members in the system because a team consists of different roles. Awareness features should ensure that team members are able to aware when other members have logged on and have been working on their team tasks, therefore one member should not be able to modify or erase something that another member is working on. Moderator and Teacher control should enable a teacher to monitor and facilitate a team's activities from time to time in order to ensure that the teamwork is under control and is carried out on the right track. When team works have been completed, they need to be stored and protected in the system and then can be retrieved and tracked online whenever needed in future. So security and file versioning and protection features should be included in the framework. These eight criteria reflect the main features of CSCL system usability defined by Brinck (2005) described in Section 2.3.2 and the requirements of CSCL system explained in Section 2.2.2.

4) *Error Tolerance* dimension ensures that a system has the capability of preventing errors, or helping with recovery from those errors that do occur (Green & Pearson, 2006; Quesenbery, 2003). Error handling is a critical component of interface as this helps free users from unnecessary errors (Green & Pearson, 2006; Shneiderman & Plaisant, 2004) so it is very important to a CSCL system which is accessed by many users in a team synchronously or asynchronously. The dimension has two criteria: Error rate and Error prevention. This study mainly looked into Error prevention rather than Error rate as developing and improving OJS usability is not the objective of this study.
5) *Universal Accessibility (Ubiquity)* dimension makes sure that a system can be used by different users including disabled users, who have various computing skills and who are from different places either in a country or in different countries. Internationalisation support is necessary if a CSCL system is used for CL globally. In this study, only one criterion – "Support different users" is defined for this dimension as evaluating the features of supporting multiple languages, disability users and overseas users are not included in this case study.
6) *Satisfaction* refers to the user's perceptions, feelings, and opinions of the product (Rubin & Chisnell, 2008). This dimension consists of five criteria: Usefulness/Functionality, Learnability/Predictability/Recognition/Memorability, Simplicity, Help/Documentation, and Aesthetic Design.

   A system must be useful to its targeted users, must be capable of helping these users complete their tasks or achieve their goals. Otherwise nobody wants to use the system because of its uselessness. Predictability and recognition will affect how much need for users memorising a system, determine if they are able to understand the system intuitively and to quickly find out what they should do by applying their previous knowledge and experience. Help and documentation and aesthetic design are essential to a system's usability. Documentation provides the necessary information about what features a system has and how to use the system, and helps users gain the confidence in using a system. If aesthetic design is pleasant and meets a user's preferences, the user will be happy and satisfied with using the system. Therefore, the combination of the five criteria will impact the user's feelings of satisfaction and if they are willing to use the system. A satisfactory system should be useful, simple, and easy to learn and use, as well as having an enjoyable aesthetic design.

Table 2-1 The Dimensions and selected Criteria in the Framework

Dimensions	Criteria	References
Effectiveness	1 Completeness, 2 Visibility, 3 Organisation/Design, 4 Navigability	(Dawson, 2006); (Folmer & Bosch, 2003); (Green & Pearson, 2006); (González, et al., 2006); (Hu & Chang, 2006); (Nielsen, 1994); (Quesenbery, 2003); (Rubin & Chisnell, 2008); (Ryu, 2007); (Sauro & Kindlund, 2005); (Scholtz, 2004); (Ssemugabi & Villiers, 2007a);
Efficiency	1 Speed, 2 Familiarity/Consistency/ Standards, 3 Flexibility/Adaptability/ Configurability, 4 Effort	(Bevan, 1995); (CUD, 2008); (Dawson, 2006); (Folmer & Bosch, 2003); (Green & Pearson, 2006); (González, et al., 2006); (Hu & Chang, 2006); (Nielsen, 1994); (Quesenbery, 2003); (Rubin & Chisnell, 2008); (Ryu, 2007); (Sauro & Kindlund, 2005); (Scholtz, 2004); (Shneiderman & Plaisant, 2004); (Ssemugabi & Villiers, 2007a)
Collaborativity	1 User Management, 2 Awareness, 3 Communication, 4 User Control/Moderator & Teacher Control, 5 File /Content Sharing /Management, 6 Process Tracking/ Automated Notification, 7 File/Content Protection, 8 Security	(Brinck, 2005); (Cyzyk & Choudhury, 2008) ; (Dawson, 2006); (Dimitracopoulou, 2005); (Graham & Misanchuk, 2004); (Kildare, et al., 2006); (Spada, et al., 2005)
Error Tolerance	1 Error Rate 2 Error Prevention	(Dawson, 2006); (Folmer & Bosch, 2003); (Green & Pearson, 2006); (González, et al., 2006); (Nielsen, 1994); (Quesenbery, 2003); (Ryu, 2007); (Sauro & Kindlund, 2005); (Scholtz, 2004); (Shneiderman & Plaisant, 2004);
Universal Accessi-bility (Ubiquity)	1 Support different users	(CUD, 2008); (Cyzyk & Choudhury, 2008); (Dawson, 2006); (Folmer & Bosch, 2003); (Rubin & Chisnell, 2008); (Shneiderman & Plaisant, 2004); (Ssemugabi & Villiers, 2007a);
Satisfaction	1 Usefulness/Functionality, 2 Learnability/Predictability /Recognition/Memorability, 3 Simplicity, 4 Help/Documentation, 5 Aesthetic Design	(Ardito, et al., 2006); (Bevan, 1995); (Cyzyk & Choudhury, 2008); (Dawson, 2006); (Dimitracopoulou, 2005); (Folmer & Bosch, 2003); (Green & Pearson, 2006); (González, et al., 2006); (Hu & Chang, 2006); (Nielsen, 1994); (Pinelle & Gutwin, 2008); (Quesenbery, 2003); (Rubin & Chisnell, 2008); (Ryu, 2007); (Sauro & Kindlund, 2005); (Scholtz, 2004); (Shneiderman & Plaisant, 2004); (Ssemugabi & Villiers, 2007a); (W3C, 2008)

2.7 Summary

This chapter has reviewed the concepts and the features of CL and CSCL system, usability and CSCL system usability, and UE and CSCL system UE, discussed some related frameworks and criteria for UE, then identified some issues and difficulties in CSCL system UE and the need for developing an effective framework for CSCL system UE.

The outcomes of the review of several related studies show that the features of a system's usability are affected by the goals of a system and its context. The three dimensions (effectiveness, efficiency, and satisfaction) defined by ISO 9241-11 standard were widely used in system usability evaluation. For groupware, some frameworks were developed and employed for UE in different contexts. But they are not capable of effectively evaluating modern CSCL online systems with new CL technologies that include real time communication, collaborative interaction, multimedia publication, security, and so on. An effective and comprehensive framework for CSCL system UE in educational settings had not been developed yet. This research had taken the opportunity and looked into this area, and proposed a framework with six dimensions for CSCL system UE in educational settings.

Chapter 3

3. Research Methodology and Design

This chapter includes five sections in the following order: defining the research method for this research, reviewing some related UE methodologies, illustrating this research methodology and the UE method design, and ending with a summary.

3.1 Research method –Case study

3.2 Review of related UE research methodologies

3.3 The methodology of this study

3.4 UE method design

3.5 Summary

3.1 Research method – Case study

3.1.1 Case study

Case study research method has been a common research strategy used in many social and science areas such as psychology, sociology, political science, social work, information systems, and so on (Yin, 2004). It is an empirical inquiry that investigates a contemporary phenomenon within its real-life context, especially when the boundaries between phenomenon and context are not clearly evident (Yin, 1994, 2003a). Case study researchers normally seek to identify themes or categories of behaviours, contemporary phenomena within their real-life contexts (Hancock & Algozzine, 2006; Yin, 1994, 2003a). Therefore Case study method is appropriate to use when:

1) the phenomenon under investigation is not readily distinguishable from its context (e.g. topics are defined broadly or not narrowly enough, a phenomenon may have complex multivariate conditions, and rely on multiple sources of evidence) or a software program may be problematic;
2) the studies are part of a cumulative body of knowledge;
3) the studies became an integral part of evaluation research and have been associated with process evolutions during the latter decades of the 20th century;
4) a study is an in-depth description and analysis of a bounded system or aims to explore a single phenomenon (the case) and to obtain in-depth knowledge in a natural setting.

(Collis & Hussey, 2009; Merriam, 2009; Yin, 2003b)

Based on the purposes of a research, Case study can be exploratory case study (to explore those situations in which the case being evaluated has no clear, single set of outcomes; to help identify questions, select measurement constructs, and develop measures), or descriptive case study (to describe a case and the real-life context in which it occurred; and to illustrate certain topics within an evaluation and answer what questions), or explanatory case study (to explain the causal links in real-life cases that are too complex for the survey or experimental strategies and to analyse and explain why or how questions) (Davey, 1991; Yin, 2003a). According to the number of cases to be studied, Case study can be designed as single-case (holistic), single-case (embedded), multiple-case (holistic), and multiple-case (embedded) (Yin, 2003a).

Figure 3-1 depicts the four stages of Case study method. Stage 3 may be back to stage 1 when it is needed (Yin, 1994, 2004).

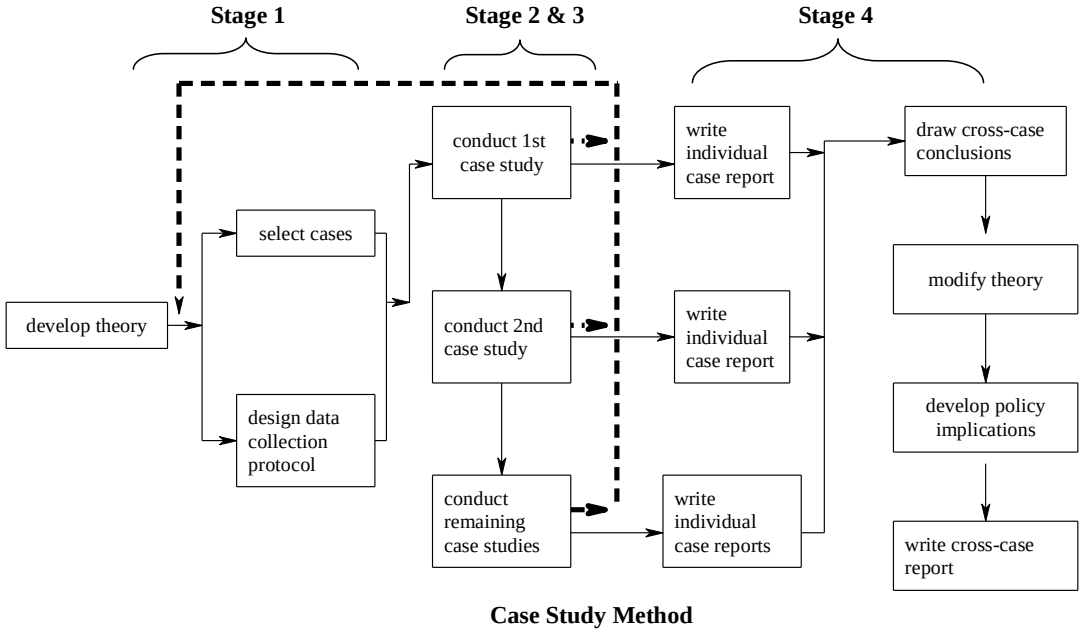
Stage 1 - Planning and Designing a case study;

Stage 2 - Conducting a case study (Preparing for data collection, collecting the evidence/data); then conducting another study as planned;

Stage 3 - Analysing evidence/data;

Stage 4 – Writing/Presenting a report.

Figure 3-1 Four Stages of Case Study Method



Sources: COSMOS Corporation. (Cited in Yin ,1994, 2003a)
(**note**: the diagram was modified by the author)

3.1.2 Why was Case study method employed in the research?

Selecting an appropriate research method must be driven by the study's area, topic, objectives, questions, and practical issues (e.g. time, money, the size of a research team, etc.) (McNeill & Chapman, 2005; Merriam, 2009). This study is a qualitative research in the area of developing a CSCL system UE framework and had adopted exploratory Case study as its research method. Three main reasons are discussed as below.

Firstly, this study aimed to identify and develop an effective framework for CSCL system UE in educational settings and to make the contribution to filling in a gap between CSCL system UE framework development in educational settings and groupware UE framework development in general settings as discussed in Sections 2.5.4 – 2.5.6. The research was planned firstly to define the criteria and dimensions to form a framework based on reviewing existing related studies in UE and CSCL usability, secondly to employ the proposed framework to evaluate one CSCL system's usability in an educational environment, i.e. in a case (a real-life context), where one CSCL system (OJS) was set up and became accessible to the potential participants (students and staff at an university), lastly to conclude if the framework would be an effective framework for CSCL system UE and to make some recommendations for the future study in developing an effective framework for CSCL system UE in educational settings.

Secondly, this study is 2 year part-time study. It was designed to look into one case – a CSCL system - OJS, used in collaborative teaching and learning environment. OJS was set up for the participants to apply the framework to do usability evaluation and also for the postgraduate students, who studied the postgraduate paper named as "Collaborative Computing" (CC), to work on their group assignments. The participants in a pilot study were recruited from academic staff, and the participants in the full-scale study were recruited from the postgraduate students in the CC class at SMCS, AUT in S2 2009.

Thirdly, the definition and explanation of Case study in Section 3.1.1 have illustrated that Case study is an integral part of evaluation research within its real-life context and relies on multiple sources of evidence (Yin, 1994, 2003a). Exploratory Case study can help identify questions, select measurement constructs, and develop measures (Davey, 1991). Therefore Case study best fits in with the goals and the situation of this study.

Grounded theory method was not selected for this study because Grounded theory aims to understand how people make sense of their experience, to identify core categories and codes which are conceptual elements of the theory, to lift data to a conceptual level,

and then to build a substantive theory grounded in the collected data. (Chamberlain, 1995; Merriam, 2009; Suddaby, 2006). Ethnography and Action research and Experiment are not the proper research methods for this study because exploring cultural patterns in the OJS community and improving OJS are not the objectives of this study. Ethnography aims "*to uncover tacit knowledge of participants in the specific culture under study, and is most likely to deal with interpersonal interaction*" (Cohen & Court, 2003) whereas Action research usually has developers and software designers involved in and intends to *"describe ongoing evaluation during which improvements are made"* (Preece, 2000). Experiment best fits in running a system where some certain settings or variables have been preset in order to test a hypothesis of a theory or some functions/features of a system (Dix, Finlay, Abowd, & Beale, 2004).

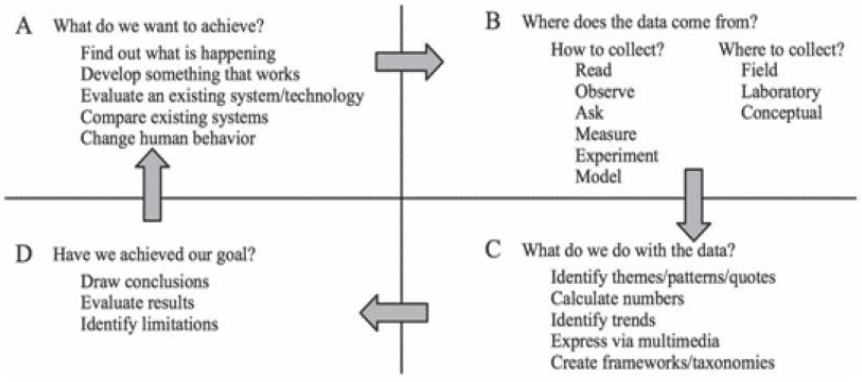
3.2 Review of related UE research methodologies

The purpose of developing UE Methodology is to provide a guideline for a research in usability evaluation. Various UE methodologies were recommended by different researchers. Some common steps of doing UE, which can be found in many of the methodologies listed in Appendix 2, are: identifying the goals of the research based on its context(s), selecting measures, deciding evaluation methods, applying the methods in a selected system, then collecting data and analysing data, and finally making conclusions and recommendations.

In this research, the following two methodologies were selected as the foundation of developing the methodology for this research because they fit well in the context of this study and are useful for defining the steps of the study in a logical sequence.

3.2.1 Computing research methods (CRM)

Figure 3-2 Methodology for computing research methods (CRM)



Source: (Holz, et al., 2006)

The methodology for Computing Research Methods (CRM) is grounded in four questions and each question anchors a quadrant in the process of computing research (Holz, et al., 2006) (as shown in Figure 3-2 above). CRM is a simple methodology for designing a research method in UE.

3.2.2 Three-step holistic approach

The Three-Step Holistic (TSH) approach consists of the following steps: (1) the development of a holistic usability framework for a system, (2) surveys of users to validate and refine the framework, and to determine attribute weights, and (3) the application of this framework to the existing system, including the development of a technique to measure holistic usability (Dawson, 2006).

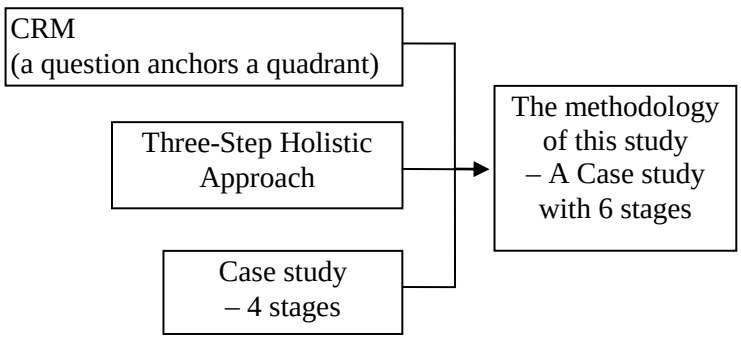
The approach is especially helpful when developing a framework for a complex system's usability evaluation as it looks not only at the user interface, but also at various aspects of system design, installation, maintenance, and use (Dawson, 2006). Therefore, this approach was selected and adapted when developing the methodology for this research.

3.3 The methodology of this study

3.3.1 The methodology of this study

Figure 3-3 shows that this research combined the two methodologies (CRM and TSH approach) and Case study research method together, and developed its own methodology.

Figure 3-3 The methodologies employed in this research



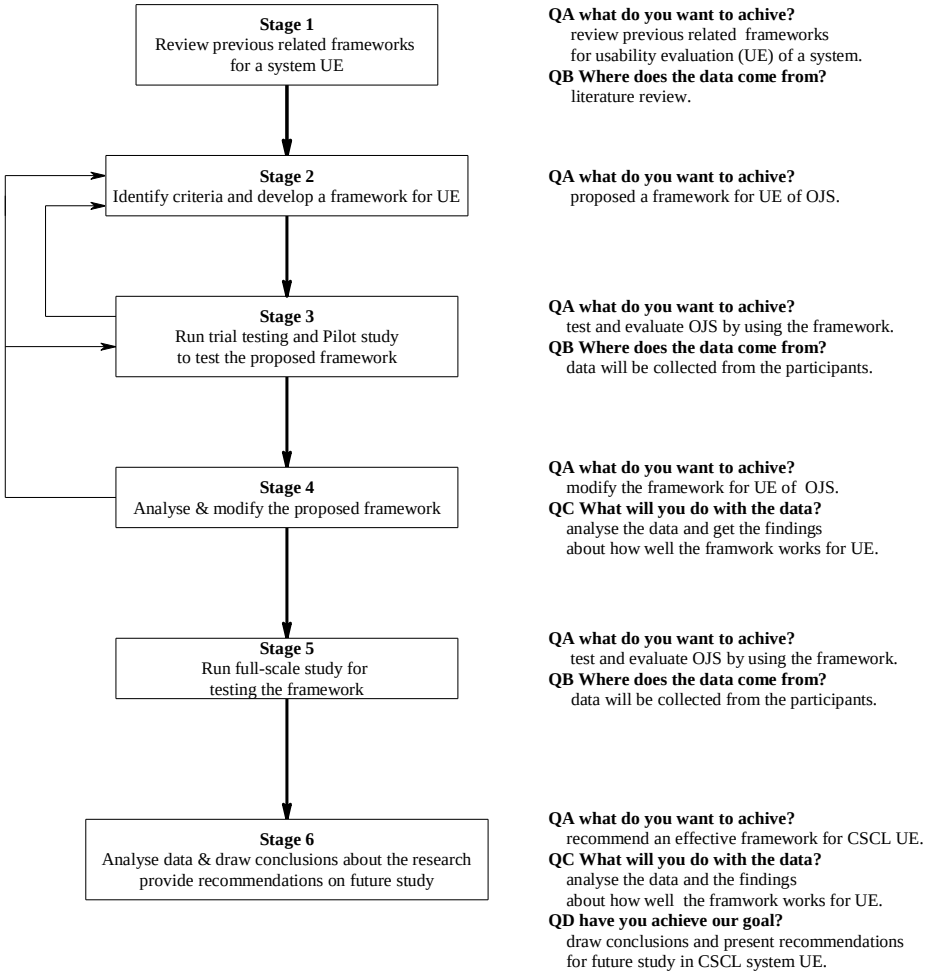
The methodology of this research is a Case study including six stages implemented in a collaborative teaching and learning environment where OJS was set up for participants to perform collaborative team work. Two or three questions from CRM were employed to define what should be achieved and how to achieve it in each stage of the study. User

testing, questionnaire, and interview were chosen as the UE methods for the study. Figure 3-4 illustrates the methodology of the study. The six stages are:

stage 1*:* set up a plan of the process of the research, review literature to examine the previous related frameworks for UE of a system, access OJS to become familiar with its features; (equivalent to step1 in TSH, stage 1 in Case study);

stage 2: analyse and compare the findings about the related frameworks and then identify the important criteria and develop a framework for UE of OJS and design a testing task; (equivalent to step1 in TSH, stage 1 in Case study);

Figure 3-4 Methodology of this research



stage 3: invite some potential participants to do trial testing and join in a pilot study group; apply the proposed framework to OJS usability testing and evaluation, collect data from OJS usability testing, the questionnaires, and the interviews; (equivalent to step2 in TSH, stage 2 in Case study);

stage 4*:* analyse the data gathered from stage 3, find out the problems or issues with the framework, the testing, and questionnaires, and make any necessary modifications; (equivalent to step2 in TSH, stage 2 in Case study)

Stages 2, 3, and 4 can be part of an iterative process where any modifications to the framework will be incorporated;

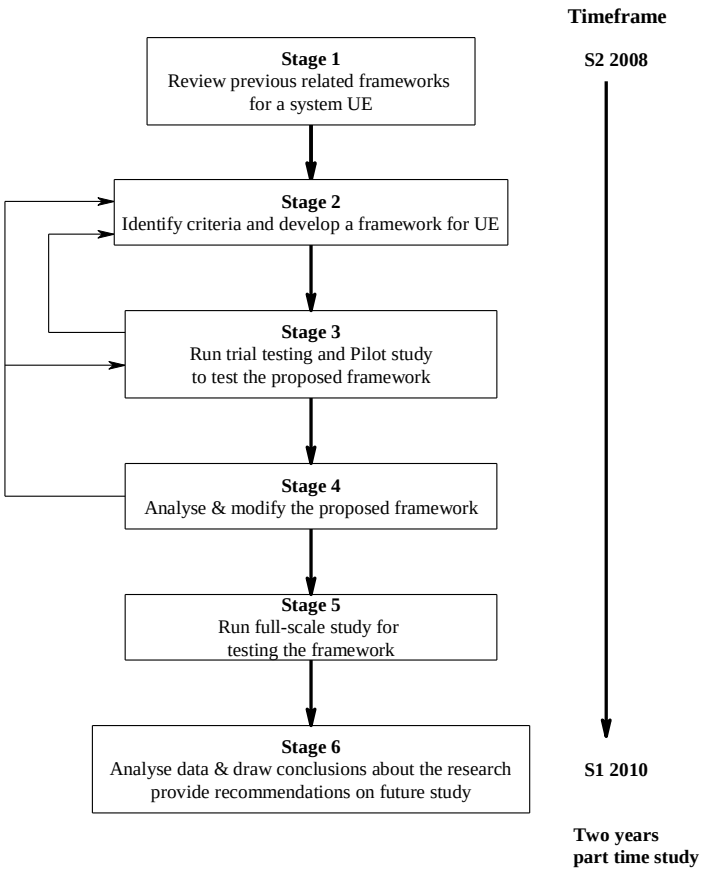
stage 5 - the full-scale study: apply the framework to the UE of OJS in the collaborative learning environment, collect data from OJS testing, the questionnaires, and the interviews; (equivalent to step3 in TSH, stage 2 in Case study)

stage 6*:* analyse data and draw conclusions, make some recommendations on what should be improved in future study in developing an effective framework for CSCL systems UE; and finally complete the research report (equivalent to step3 in TSH, stage 3 & 4 in case study).

3.3.2 Research Planning

Figure 3-5 shows that this research is a part time study and was planned to take two years from Semester 2 (S2) 2008 to Semester 1 (S1) 2010.

Figure 3-5 The planned timeframe of this research



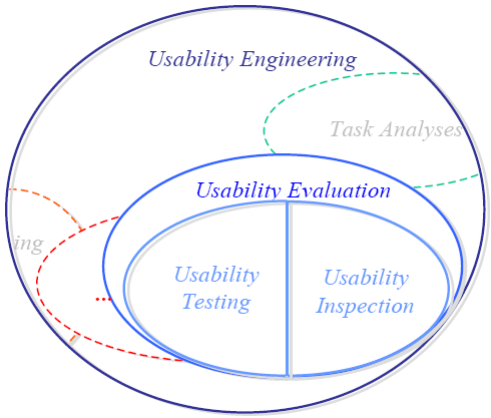
3.4 UE method design

This section will review UE methods and explain the UE methods defined by the methodology of this research.

3.4.1 Review of UE Methods

Various UE methods can be found in the previous related researches and publications and can be categorised in different ways. Gellner and Forbrig (2003) defined two types of UE methods - inspection and testing (see Figure 3-6) while Zhang (2008) added one more type and divided UE methods into three types: testing, inspection, and inquiry. Inquiry methods include interview, questionnaire, field observation, and so on.

Figure 3-6 Layer of Usability Terms



(Gellner & Forbrig, 2003)

According to the type of data collected, UE methods can be *qualitative* or *quantitative* whereas UE methods can be *formative* or *summative* based on the stage of the lifecycle of a system/product at which a UE is carried out. *Formative* evaluations are used to obtain information used in design (Scholtz, 2004), and conducted during the design and construction phase. Its results and conclusions are used mainly for bug fixing and improving the characteristics of the interface and providing input for redesign in order to improve usability (Karoulis & Pombortsis, 2003; Zhang, 2007). In contrast, *summative* evaluations are usability evaluations that document the effectiveness, efficiency, and user satisfaction of a product at the end of the development cycle (Scholtz, 2004), and conducted after the product has reached the end user. The results and conclusions are used to improve the interface as a whole and meet more user needs in a following upgrade (Karoulis & Pombortsis, 2003). Typical summative usability measures include task success rate, task completion time, error rate, subjective satisfaction rating, and so forth (Zhang, 2007).

UE methods can be *User-based methods/Direct methods/Empirical* evaluations or *Expert-based methods/Indirect methods/Usability inspection* depending on who performs the evaluation. *User-based methods/Direct methods/Empirical evaluations* mainly consist of user testing. The evaluations are accomplished by identifying

representative users, representative tasks, and developing a procedure for capturing the problems on a tested system (Nogueira & Garcia, 2003; Scholtz, 2004) while *Expert-based methods/Indirect methods/Usability inspection* are an efficient formative evaluations method and can be applied even on system prototypes or design specifications up to the almost ready-to-ship product (Karoulis & Pombortsis, 2003; Scholtz, 2004). Only expert evaluators inspect the application and provide judgments based on their knowledge and experience (Molich & Dumas, 2008; Zhang, 2007). The experts can be software developers, usability engineers, and other professionals who are deeply familiar with the concepts of usability in design. The popular inspection methods are Cognitive Walkthroughs, Feature Inspection, Heuristic Evaluation, Pluralistic Walkthrough, Perspective-based Inspection (Zhang, 2008).

In this research, user testing was adopted because the research did not aim to improve or develop a system itself but to test a framework, i.e. to use the framework to evaluate the system's usability and then to find out if it is capable of evaluating a CSCL system's usability and identifying advantages and disadvantages of the system usability. Therefore, the UE method in this research as a whole is a qualitative, summative, user-based usability testing, and consists of user testing, questionnaire, and personal interview. Think Aloud and user observation are the techniques of usability testing. User testing, questionnaire, and interview were undertaken during a testing session.

3.4.1.1 User testing

User testing has a long history (Downey, 2007) and gained popularity in the early 1980s (Dumas & Fox, 2008). It is widely used in usability evaluation (Nakamichi, Shima, Sakai, & Matsumoto, 2006). It is "*a process that employs people as participants who are representative of the target audience to evaluate the degree to which a product meets specific usability criteria*" (Rubin & Chisnell, 2008). During user testing, participants work on typical tasks using the system (or the prototype) and are observed; the evaluators use the result to see how the system supports the users to do their tasks (Zhang, 2007) and to evaluate the degree to which a system/product meets specific usability criteria (Rubin & Chisnell, 2008). It can often uncover very specific areas needing improvement (ForakerDesign, 2005). About the sample size of a research, Rubin and Chisnell (2008) advised that "*for the purpose of conducting a less formal usability test, research has shown that four to five participants who represent one audience cell will expose about 80 percent of the usability deficiencies of a product for that audience, and that this 80% will represent most of the major problems*". However

Hwang and Salvendy (2010) suggested that a general rule for optimal sample size would be '10±2' instead of '4±1'. For example, Think Aloud requires nine test users to reach 80% overall discovery rate.

Table 3-1 shows that usability testing (user testing) method has different techniques, such as Coaching Method, Co-discovery Learning, Question-asking Protocol, Remote Testing, Think Aloud Protocol, User Observation, and so on (Gomoll & Nicol, 1990; Ivory & Hearst, 2001; Zhang, 2007, 2008). In this study, Thinking-Aloud Protocol and User observation were employed.

Table 3-1 Usability Testing Techniques

Usability Testing Techniques	**Description**
Thinking-Aloud Protocol	user talks during test
Question-Asking Protocol	tester asks user questions
Shadowing Method	expert explains user actions to tester
Coaching Method	user can ask an expert questions
Teaching Method	expert user teaches novice user
Co-discovery Learning	two users collaborate
Performance Measurement	tester records usage data during test
Log File Analysis	tester analyzes usage data
Retrospective Testing	tester reviews videotape with user
Remote Testing	tester and user are not collocated during test
User observation	Observe watches and listens carefully to users as they work with a product or a system

(Gomoll & Nicol, 1990; Ivory & Hearst, 2001)

Thinking-Aloud Protocol

Think aloud protocol is commonly employed with usability testing and evaluation (Molich & Dumas, 2008; Hwang & Salvendy, 2010). It is "*a simple technique intended to capture what the participants are thinking while working*" (Rubin & Chisnell, 2008) and *"valuable because of the rich qualitative data it produces. It helps identify areas of the product that could be improved"* (Dumas & Loring, 2008). During a testing session, the participants are asked to verbalise their thoughts, feelings, and opinions while interacting with the system (Rubin, 1994). With permission of the users, what they say is recorded and analysed for explanations of what they are trying to do; how they interpret the items shown on the computer screen; and explanations of any difficulties that they may have (Smith-Atakan, 2006). This technique offers a window over the users' mental models, and allows evaluators to detect possible misconceptions about the system and the interface elements which cause them (Ardito, et al., 2006). It not only is useful in capturing a wide range of cognitive activities and best suitable for verifying

ease of use and ease of learning (Nogueira & Garcia, 2003; Zhang, 2007) but also is a cheap way of getting a lot of good qualitative feedback during testing (Hom).

User Observation

Observation is the most frequently used evaluation technique as per Pinelle and Gutwin's advice in their study (2000). There are two observation methods for usability testing. '*Unobtrusive observation' means you observe what test users do and refrain from interacting with them. 'Obtrusive observation' means you interact with test users*" (D'Hertefelt, 1999). The technique used in this study is unobtrusive user observation.

Gomoll and Nicol (1990) recommended a guideline containing 10 steps for conducting a user observation: *"Introduce yourself; describe the purpose of the observation; tell the participant that it's okay to quit at any time; talk about the equipment in the room; explain that you cannot provide help; describe the tasks and introduce the product; ask if there are any questions before starting the testing; conclude the observation when the test is over; and use the results".* For unobtrusive observation, D'Hertefelt (1999) advised that an observer should "*be quiet and just watch, first observe then take notes, stimulate users to think aloud, only help to overcome the limitations of the prototype, elicit detailed information, answer test user questions with questions, and limit the time test users have to execute a task".*

According to the guideline and the advice above, during a testing session, the researcher acted as the facilitator and the observer, gave a participant a brief introduction about the purpose of the testing and what needed to be done; then sit near to a participant quietly and watched what a participant was doing, stimulated the participant to think aloud, recorded the time of starting and completing the task and what questions were asked and what errors occurred on the Observer data collection sheet (ODCS) (see Appendix 17) while observing. When the participant got stuck, the facilitator/observer asked questions to help the participant think in different ways in order to avoid running out the time and to ensure that the testing was under control.

3.4.1.2 Interview

Interviews can be structured or unstructured. "*Unstructured Interviews are used during earlier stages of usability evaluation. ... The evaluator does not have a well-defined agenda and is not concerned with any specific aspects of the system. A structured interview has a predetermined agenda which specific questions to guide and direct the*

interview. … with permission of the user, the interview session should be recorded" (Zhang, 2007). Structured interviews are useful for comparing the responses of different interviewees but have a disadvantage -– the evaluator is not allowed to ask follow-up questions. Unstructured interviews overcome the disadvantage of structured interviews but it can be difficult for the evaluator to collate and analyse the obtained data (Smith-Atakan, 2006).

This research used structured interview by asking some prepared close-ended and open-ended questions so the data collected from different participants could be compared and analysed.

3.4.1.3 Questionnaire

Questionnaire is a written list of questions that are distributed to users (Hom). It is best suitable for verifying user satisfaction and productivity. Participants just need to answer the questions (Nogueira & Garcia, 2003). The biggest advantage is *"that a usability questionnaire gives you the feedback from the point of view of the user. … Questionnaires are usually quick and therefore cost effective to administer and to score*" whereas the biggest disadvantage is *"that a questionnaire tells you only the user's reaction as the user perceives the situation"* (Kirakowski, 2000). Some responses may be cheating because participants just simply give the same answer to all the questions without considering each question carefully.

In this research, pre-test questionnaire and post-test questionnaire were given to the participants. The former consists of 14 multiple choice questions - four are demographic questions and ten are about participant's experience (see Appendix 3) and intends to gather information which determines which user type a participant belongs to. The latter was developed according to the proposed framework and its criteria, and consists of three parts: Likert-scale questions for ranking agreement scales on the statements about the criteria related to the features of OJS in part 1, Likert-scale questions for ranking levels of the importance of the criteria to a system UE in part 2, some open-ended questions about the framework and OJS usability in part 3 (see Appendix 4).

3.4.1.4 User types in usability testing

A system is normally developed and used by a variety of users with different skills. Users are the people who use the system to achieve their goals. Persona is a term for user type or user profile and referred as "*specific people with particular needs, wants,*

desires, skill levels, and contexts of use. … *The essential characteristics of a persona include: a name and a picture, demographic information (age, education, ethnicity, family status), job title or main focus of activity ((*e.g.*) student ...), goals (product related and experience related), environment (context of use ...), technical or product domain expertise, and a quote that sums up what matters most to the persona"* (Barnum, 2011). Domain knowledge and technical skills mainly include familiarity with the system, computer and software skills, the usages of the system, and users' attitudes toward technology (Barnum, 2011).

For examining user's perception towards e-learning system usability, Koohang (2004) recommended four variables about user type to be considered. They are age, gender, prior experience with the Internet, and the amount of time a student spent on a system. His research concluded that user's prior experience with the Internet and the amount of time he/she spent on the e-learning system were significant factors but age and gender were not. About Internet skills, Deursen and Van Dijk (2010) proposed a framework which consists of four types of skills: operational, formal, information, and strategic skills. They suggested that computer-mediated communication skills should be taken into account in future. These skills are necessary for digital environment, specifically the Internet, and reflect user individual's ability of using online system.

"*Usability is the ability of a product to satisfy the needs and specifications of users*" (Koohang, 2004). For usability testing, "*selecting participants involves identifying and describing the relevant behaviour, skills, and knowledge of the person(s) who will use your product*" and the results of testing will only be valid if participants are typical users of a system, or as close to the identified users as possible (Rubin & Chisnell, 2008). So, before recruiting participants, a researcher should identify who are the users of the system and categorise personas (user types) based on the goals of the system and the study, and the skills and the characteristics of the users. A researcher should also understand that "*participants are rarely fully representative of the target population*" (Rubin & Chisnell, 2008), consider the constraints of time, money, resources, etc. and then decide how many types (personas) and how many participants in each type to be recruited in a study (Rubin & Chisnell, 2008; Barnum, 2011).

For CSCL system UE, Tselios, Avouris, and Komis (2008) pointed out that the processes "*engage students into higher conceptual activities and reflections through*

direct manipulation of entities of the learning subject, and at the same time offer adequate collaboration and communication mechanisms". Quite generally, the users of using a CSCL system "*may be students, tutors, teachers and possibly technical experts*" (Lund, 2004). They can be grouped into experts, intermediates, or novices (Barnum, 2011). In this study, OJS was set up and became available to the postgraduate students in SCMS at AUT for the period when the postgraduate paper - Collaborative Computing (CC) was taught. It was used as a CL system for peer review group assignments and sharing and storing the students' opinions and information. Only the postgraduate students in CC class could access the system for the certain period. A few lecturers in SCMS used OJS as a journal publishing system which was located in a different sever. So, the existing users of OJS were the students and the lecturer in CC class during the period of the study and could be recruited as the participants of this study. Other potential participants could be the postgraduate students, who had completed the paper or had not done the paper, and the staff, who might be interested in this study and had used or had not used OJS, in SCMS at AUT. Both the staff and postgraduate students had worked on computers and accessed the Internet and online systems for years. The details about participant recruitment are presented on Section 4.4 pg58.

This study considered the suggestions on user categorisation from Koohang (2004) (four variables), Barnum (2011) (domain knowledge and technical skills), and Deursen and Van Dijk (2010) (five types of skills), and divided the users/participants of OJS usability testing into five groups i.e. novice, beginner, intermediate, advanced, and expert mainly based on user's prior experience in using CSCL system and the amount of time spent on OJS (see Table 3-2). Pre-test questionnaire was created for getting this information (see Appendix 3).

Table 3-2 User Types

User type	Novice	Beginner	Intermediate	Advanced	Expert
The length of time of having used CSCL system	N/A	≤ 1 yr	1-2 yrs	2-3 yrs	≥ 3yrs
Time spent on using OJS per week	Null	≤ 1hr	1hr - ≤ 4hrs	4hrs - ≤ 10hrs	≥ 10hrs

3.4.1.5 The Roles of the researcher

A usability test team normally consists of different roles, such as facilitator, user, observer, and system developer (Snyder, 2004; Spool, 2009). Depending on the type of a system to be tested and the number of people are involved, combining roles may become necessary. "*A typical test session involves one participant and one moderator,*

lasts one to two hours" (Dumas & Loring, 2008). So, a testing team has at least two roles, i.e. facilitator (or called as moderator) and participant (or called as user). A researcher can be a facilitator and also be an observer. An observer usually remains silent and takes notes while he/she observes what the participants are working on the tasks during a testing session (Snyder, 2004).

The responsibilities of a facilitator include preparing testing documents, task scripts and checklist before a testing, greeting the participants and giving them pretest briefing and obtaining informed consent from them at the beginning of the testing; monitoring the participants throughout a set of selected tasks of using a system during the testing; and debriefing the observers at the end of the testing if needed (Snyder, 2004). During a testing, the participants should complete tasks on their own with little help. However a facilitator can give assistance when the participants are stuck, and there is more to be learned if they continue, or when they have unknowingly gone too far down the wrong path and need to be brought back on track, or when they should move on to the next task because the problem they are having are not new and/or time is limited (Dumas & Loring, 2008). When Think-aloud protocol is employed in a usability testing, a facilitator normally sit close to the participants, encourage them articulate their thoughts while they are working on the tasks, ask questions but remain neutral.

Because the ways that a facilitator interacts with the participants have *"a huge effect on test results*" (Dumas & Loring, 2008), Snyder (2004) suggested that a facilitator should act "*like a duck --serene on the surface, but paddling like heck underneath"* and Spool (2009) believed that the best facilitators "*have a lot in common with an orchestra conductor*". So, a facilitator should have three roles - *Flight attendant* (focusing on safety and comfort, ensuring that participants do not have an unpleasant experience), *Sportscaster* (catching all the actions, and maximising the flow of information from the participants to the observers), and *Scientist* (guiding the data collection, maintaining the highest possible degree of integrity in the data) (Snyder, 2004; Spool, 2009). Similarly, Dumas and Loring (2008) advised that a facilitator's roles should be the Gracious Host, the Leader, and the Neutral Observer when moderating a testing. They recommended 10 golden rules for a facilitator as listed below:

1) *Decide how to interact based on the purpose of the test.*
2) *Respect the participants' rights.*
3) *Remember your responsibility to future users.*

4) *Respect the participants as experts, but remain in charge.*
5) *Be professional, which includes being genuine.*
6) *Let the participants speak.*
7) *Remember that your intuition can hurt and help you.*
8) *Be unbiased.*
9) *Don't give away information inadvertently.*
10) *Watch yourself to keep sharp*.

Two roles of people were involved in the user testing in this study. One is facilitator and the other is participant. Two persons came to each testing session, i.e. the researcher acted as the facilitator and observer and a participant as well, and one participant worked on the defined tasks for testing OJS. Because "*Taking notes while listening can be a very difficult skill to master*" (Dumas & Loring, 2008), the facilitator used a voice recorder and screen capture to record what a participant had said and done in order to get all the data about the system usability and the testing. 10 golden rules above were adopted as a guideline for moderating the testing sessions in this study. The details on user testing design are illustrated on next section.

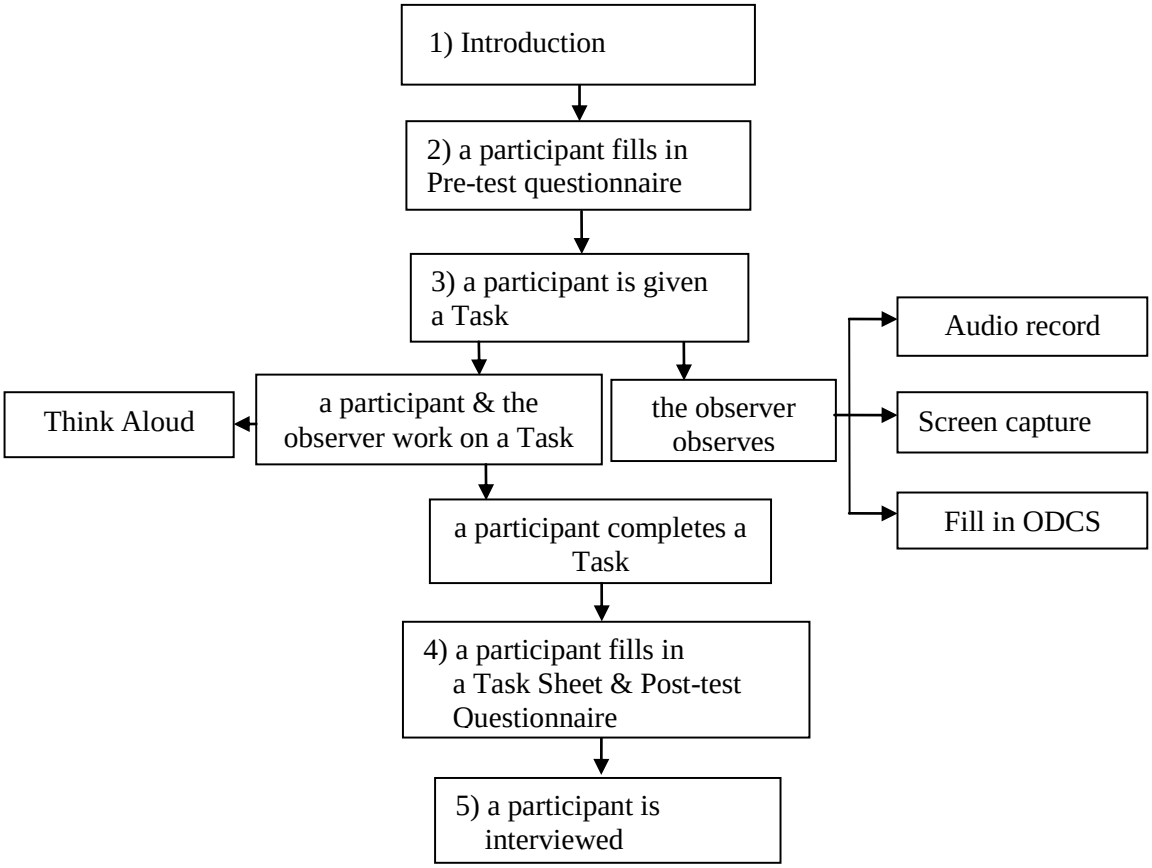
3.4.2 User testing design for the study

The time length of a testing session was set to 60 minutes. The researcher was the facilitator and also acted as the observer and a participant in each testing session. Figure 3-7 next page shows that a testing session consists of the following five steps:

1) a participant was given a brief introduction and a task script which shows the steps of doing a collaborative task. The facilitator needed to get the permission from the participant before turning on a digital voice recorder and running screen capture on the computer where the participant was working.
2) the participant was required to complete a pre-test questionnaire.
3) the participant was required to Think Aloud while working on the testing task. When a participant was doing the testing, he/she had to speak loudly so the observer would be able to know what his/her questions, comments, and thoughts about OJS usability were. The participant and the observer worked on the task collaboratively. Meantime, the observer observed and made notes on ODCS. A digital voice recorder was turned on to record the participant's comments. Camtasia Studio captured the screen activities on the computer that was used by the participant. The time allocated to the task was about 20 minutes.

4) the participant needed to complete a task sheet (TS) and a post-test questionnaire and then hand them back to the observer.

Figure 3-7 User testing Design
(up to 60 minutes**)**



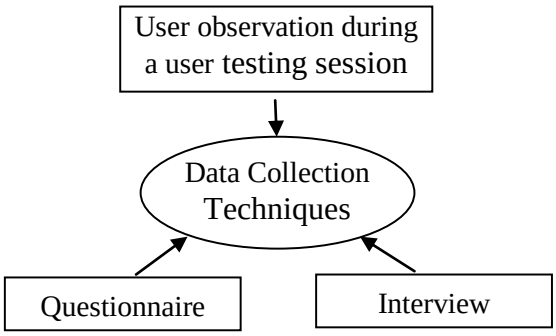
5) the participant was interviewed. 10 predefined questions (see Appendix 5) were asked. The questions focused on the feedbacks about OJS usability, advantages and disadvantages of OJS system for collaborative teamwork, the importance of the criteria and six dimensions in the framework to CSCL system UE, and the suggestions on the testing. A digital voice recorder was turned on to record the talk during the interview.

The collaborative testing task in this study will be described in Section 4.2.2.

3.4.3 Data Collection Techniques

Techniques of data collection in this research include user observation, questionnaire, and interview (Figure 3-8 below).

Figure 3-8 Various Data Collection Techniques



As explained in Section 3.3.1 (pg33) and last Section, user observation was undertaken during user testing on the stage three - pilot study and the stage five – full-scale study. The researcher was the observer who filled in the ODCS to mainly record the time spent on the task, the errors, and questions asked by participants. Questionnaires were given to the participants and Interview were conducted at the stages three and five, i.e. during a testing session a participant was required to complete a TS, a pre-test questionnaire, a post-test questionnaire, then was interviewed at the end of the testing session. Questionnaires were also given to the participants in the follow-up group.

So, the data on this research were collected at stage three and five from multiple sources, i.e. user observation, the questionnaires, and interviews. Most data is qualitative data in multiple forms such as text, digital audio and video. Some quantitative data were gathered from some questions in the questionnaires.

3.5 Summary

This chapter has reviewed the existing methodologies, Case study research method and UE methods. CRM and TSH approach and Case study research method were synthesised to develop the methodology for this research. The methodology defines that this research is an exploratory Case study that has a single case and includes six stages. The UE methods consist of user testing (Think Aloud protocol and user observation), Interview, and Questionnaire. Five user types were categorised. The researcher and one participant attended each testing session. The roles of the researcher in this study were a facilitator, an observer, and a participant. The research was a part-time study, planned to take two years and to collect qualitative data and quantitative data by employing various data collection techniques such as user observation, questionnaire, and interview.

Chapter 4

4. Case Study

This chapter consists of nine sections as listed below. It introduces OJS and defines its role in this study, illustrates the testing task design, explains the trial testing and the modifications made on the framework, questionnaires and the task script, and describes the stages of participant recruitment. Then it gives details on what had been done in the pilot study and the full-scale study and what data had been collected and how the raw data was processed.

4.1 Open Journal Systems (OJS)

4.2 Testing task design

4.3 OJS set up & Trial testing

4.4 Participant recruitment

4.5 Pilot study

4.6 Full-scale study

4.7 Data collection

4.8 Data processing

4.9 Summary

4.1 Open Journal Systems (OJS)

OJS was set up and used as a CL system in this study.

4.1.1 Introduction of OJS

OJS is open source software released under the GNU General Public License[2] and developed by the Public Knowledge Project (PKP)**.** The first PKP software was launched in 2001 (PKP, 2010). It is part of PKP's suite which consists of four separate but inter-related components - Open Journal Systems, Open Conference Systems (OCS), Open Archives Harvester and Lemon8-XML. All of the products are open source and freely available to anyone. They share similar technical requirements (PHP, MySQL, Apache or Microsoft IIS 6, and a Linux, BSD, Solaris, Mac OSX, or Windows operating system) and need only a minimal level of technical expertise to set up and run (Wikipedia.org, 2008b).

[2] **Type of open-source license:** GNU General Public License 2+ (Cyzyk & Choudhury, 2008)

OJS was designed to facilitate the development of open access, peer-reviewed publishing, providing a network where journal articles can be submitted, reviewed, indexed, and found on both a local intranet and the Internet. Participants can be assigned different roles, such as the Journal manager, editor, reviewer, author, reader, and so on (Wikipedia.org, 2008a; Willinsky, 2005). The software can be free downloaded from the Public Knowledge web site and installed on a web server which has Linux or Windows or Unix operating system, and runs Apache, PHP, and MySQL database (Willinsky, 2005). The version of OJS used in this study is 2.2.2 released on 26th Aug 2008 (PKP, 2008b) and was set up on a Linux web server located in SCMS at AUT.

4.1.1.1 The goals of OJS development

The initial goal of the system development was to create open source software that was specifically developed to manage and publish journals online and to increase open access to scholarly research and output with low cost (Willinsky, 2005). Nowadays the goal of system development has been extended. The current system aims to offer scholars a way of depositing their work, to support their involvement in the running of journals and conferences, to support a community where scholars can build their expertise in content acquisition, editorial processes, and electronic distribution, and share knowledge and studies (PKP, 2010).

4.1.1.2 Features of OJS

OJS is an online journal publishing system. It offers PDF searching, a complete help manual, multiple rounds of reviewing, automated reminders, reviewer ratings, and a host of other features. It enables a single editor to manage publishing and index peer-reviewed journals over the Internet, and also supports an international team of editors, with shared responsibilities for a journal's multiple sections (Case & John, 2007; Muthayan, 2003; Willinsky, 2005).

OJS is a text-production oriented and asynchronous system, and a highly flexible editor-operated journal management and publishing system. After the system is set up, it can be used to generate any number of journals from that site and is ready to be configured by the journal manager or editor who can do this by simply filling in a series of templates in the Setup section of the journal. The templates on the system include the journal's basic details (e.g. title of the journal; principal contact; sections of the journal, etc.) and email links, as well as provide places to post and manage journal policies, processes, and guidelines (Willinsky, 2005). The templates can be customised to meet

user's requirements. Reading Tools offer the access to related studies, media stories, government policies, etc. in open access databases (Wikipedia.org, 2008a).

OJS has a 'plugin' architecture, which allows new features to be easily integrated into the system without the need to change the entire core code base. It has various plugins, such as a tool to facilitate indexing in Google Scholar and PubMed Central, a feed plugin providing RSS/Atom web syndication feeds, a COUNTER plugin allowing COUNTER-compliant statistics and reporting, and so on. It supports Lots of Copies Keep Stuff Safe (LOCKSS[3]) that helps ensure permanent archiving for ongoing access to the content of the journal but does not support file sharing (Wikipedia.org, 2008a; Willinsky, 2005). Moreover, OJS is a multilingual system, allowing journals to publish in a variety of languages. A growing body of publications and documentation is available on the project web site. All translations are created and maintained by the OJS user community (PKP, 2010; Wikipedia.org, 2008a).

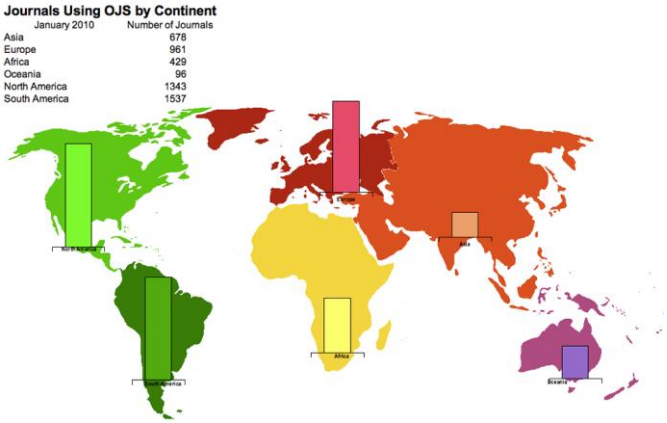
Case and John (2007) recommended OJS as the most attractive option for publishing journals in their university due to its sophisticated functionality, easy-to-use, low requirement for the expertise of current staff, and cost-effective platform. Cyzyk (2007) studied and evaluated top 7 open access journal publishing systems, and commented that OJS was his preferred journal system.

4.1.1.3 The demand on OJS in the world

PKP's suite is free, open source software for the management, publishing, and indexing of journals and conferences. Therefore, many institutions have seen the value of the PKP software, used OJS and OCS to publish their research results and the proceedings and to organise their conferences (Wikipedia.org, 2008b). The Public Knowledge Project team has seen a tremendous growth of the demand on OJS since 2005. In 2006, there were approximately 400 journals using OJS, 50 conferences using OCS, and 350 members registered on the online support forum. As of Jan 2010, over 5000 titles of Journals have been using OJS (see Figure 4-1 below) (PKP, 2010; Wikipedia.org, 2008b).

[3] The **LOCKSS** (Lots of Copies Keep Stuff Safe) project, under the auspices of Stanford University, develops and supports an open source system allowing libraries to collect, preserve and provide their readers with access to material published on the Web. The system attempts to replicate the way libraries do this for material published on paper. It was originally designed for scholarly journals, but is now also used for a range of other materials. (Wikipedia.org, 2008a)

Figure 4-1 Journals using OJS by Continent



(PKP, 2010)

4.1.2 The role of OJS in this case study

OJS was employed as a collaborative teaching and learning system for the course of Collaborative Computing (CC) in SCMS at AUT in 2008 and 2009 respectively. According to the CC course outline, the students studying CC were required to submit their group assignments in OJS, peer review their group assignments, then modify and complete the group assignments. The purposes of using OJS were different from the original purposes of developing OJS as described in Section 4.1.1.1.

In general, peer review is "*a process of quality control in which submitted papers are evaluated by experts in the field and if approved papers are published. If not approved, articles need to be corrected or can be rejected completely. Web publishing systems implement a workflow to facilitate this process*". The process "*starts at the submission of the paper and usually ends with the publication of the paper in one of the issues of the journal. … Editors and editorial staffs interact with the system assigning reviewers and preparing the issue based on their feedback"* (Chýla, 2007). The original goal of peer review in OJS is to control the quality of submissions before publishing them.

However in the CC course, students worked in teams and used the tools and virtual space provided by OJS to complete peer review processes, to post their comments and suggestions for other groups, and then to help other groups improve the quality of their assignments. By collaboratively working together, team members not only shared knowledge and learnt from one another, but also improved the skills of time management and team collaboration. So the purposes of peer reviewing in OJS in the course were for students to collaboratively work and learn together and create high quality assignments.

Therefore, the role of OJS in this case study was being a CSCL system that supported asynchronous collaborative teamwork in an educational setting. The purpose of using OJS in this study was not to publish students' assignments or provide teaching materials but to provide participants with a collaborative learning and working environment. Then the framework for CSCL system UE proposed at Stage 2 could be tested when it was applied to evaluate OJS's usability.

4.2 Testing task design

4.2.1 OJS task analysis

According to Groupware Walkthrough and Collaboration Usability Analysis (CUA) recommended by Pinelle and Gutwin (2002, 2008), "*the mechanics of collaboration are a set of group work primitives that represent the basic operations of teamwork*" and collaboration involves *taskwork* (the actions that must be taken to complete the task) and *teamwork* (the actions that must be carried out in order to complete a task as a group). This research had analysed the tasks carried out in OJS and defined five stages of the journal publishing process in OJS. The workflows of peer review, editing, and publishing are shown on Figure 4-2 on pg52. Some team tasks (Collaborative tasks) and single user tasks are listed on Appendix 7, Appendix 8, and Appendix 9.

The five stages from submitting to publishing in OJS are:

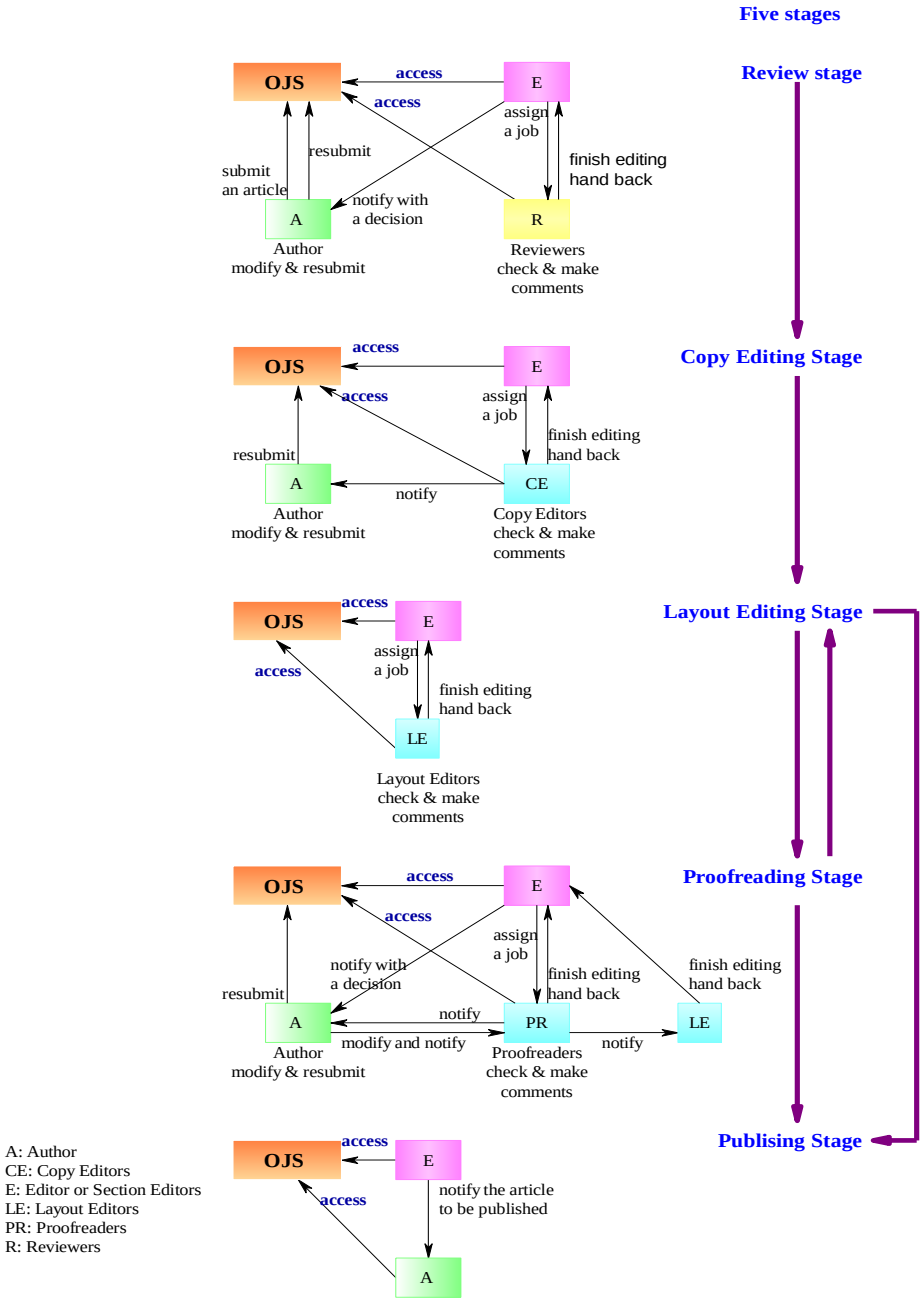
1) *Reviewing stage*: an author submits a article in OJS; an Editor/Section Editor assigns reviewers and sends the reviewers an email for review request; Reviewers accept and then review and make some comments and requirements on revision, and post them in OJS and notify the Editor with the comments; the author receives the request for revision from the Editor, modifies and then uploads the article and sends an email to the Editor; Reviewers are requested to review it, and then notify the Editor that the review has been completed and what their recommendations are; the Editor decides to accept the submission and assigns a Copyeditor and also notifies the author about the decision; but if the Editor decides to reject the submission, then he/she will notify the author with the decision. No further actions will be taken for this submission.
2) Copy editing stage: the Editor assigns a Copyeditor and sends the Copyeditor an email for editing request; the Copyeditor accepts the request, and then checks and makes some comments and requirements for modification and notifies the author about the comments/request; the author modifies and then uploads the article and

sends an email to the Copyeditor; the Copyeditor checks and accepts it, and notifies the Editor about the completion of copy editing;

3) Layout Editing stage: the Editor assigns a Layouteditor and sends the Layouteditor an email for request; the Layouteditor checks and modifies the layout of the article, then notifies the Editor that the article is ready for proofreading;
4) Proofreading stage: the Editor assigns a Proofreader and sends the Proofreader an email for Request; the Proofreader accepts the request, checks it, and posts comments/requirements for modification in OJS and also sends them to the author for revision; the author modifies and uploads the file and sends an email to the Proofreader; the Proofreader checks and accepts the article, then notifies the Layouteditor that the article is ready for checking; the Layouteditor corrects layout errors and then uploads the final version in OJS, and then notifies the Editor that the file is ready for publishing;
5) Publishing stage: the Editor makes the decision – accepts the file, schedules the article to be published and notifies author about publishing.

The tasks to be completed at each stage are collaborative team tasks. Different roles, such as an Editor, Reviewers, an Author, a Copyeditor, a Layouteditor, and a Proofreader, have to work collaboratively and effectively in order to complete each stage smoothly and get a submission published on time. The whole process in OJS is asynchronous and different roles have to work in a sequence. If one task or one stage gets stuck, the publishing process will stop at that point and cannot move on to next task or stage until it is completed.

Figure 4-2 Collaborative Teamwork Workflow of E-Journal Publishing Process on OJS



Reference: PKP, 2008a & Appendix 6

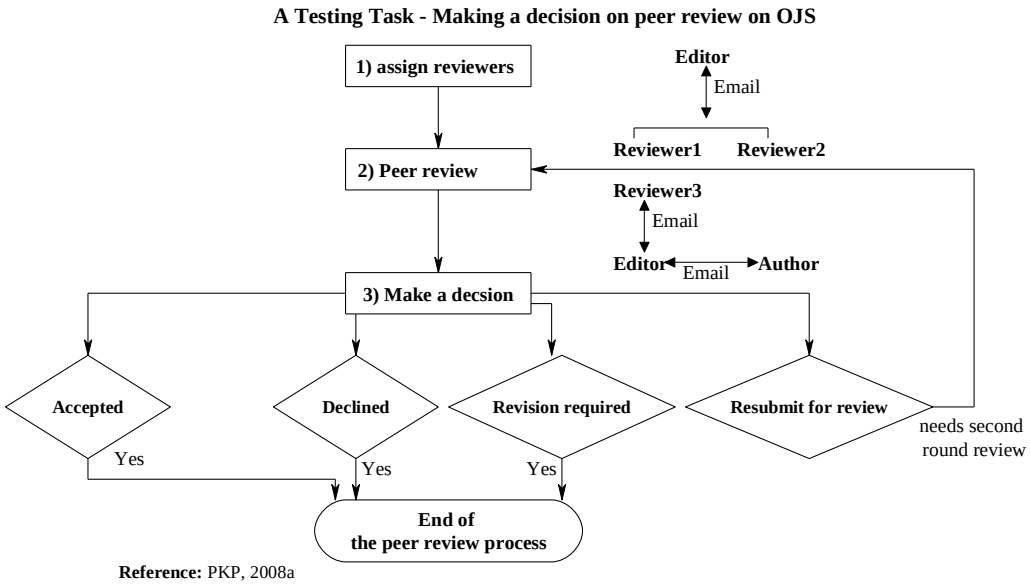
4.2.2 Testing task design

Three factors needed to be considered when designing the task for usability testing. The first factor was the time allocated to the task. It had to be completed within 20 minutes as the time span (about 20 minutes) had been defined in section 3.4.2. The second factor was the type of the task. It had to be a collaborative team task because the features of the CL system usability were evaluated. The last factor was the purpose of using OJS in the CC course because the potential participants were mainly recruited from the class. As defined in Section 4.1.2, the role of OJS in the CC course was being a CL system where students could collaboratively learn together, work on peer review and then

complete their group assignments. So, the testing task needed to be related to peer review process and could be completed around 20 minutes, and should have more than one person involved.

During a peer review process on OJS, several tasks are carried out such as enrolling users, assigning roles, submitting an article, reviewing a submission, posting comments, making a decision and notifying users, and so on (Appendix 7 and Appendix 8 show more details about OJS tasks). Users communicate with one another asynchronously via email. Initially the testing task was defined as "*Making a decision on a submission at peer review stage*". The testing task consisted of five subtasks: 1) the Journal manager or the Editor enrols users, 2) an Author submits an article, 3) the Editor assigns reviewers, 4) Reviewers review the article and post comments, and then 5) the Editor makes a decision and notifies author. At least four users would need be involved: one Journal manager or one editor, one author, and two or three reviewers. After the trial testing, the subtask - submitting an article was removed and the testing task script was modified (see Section 4.3.2 for more details).

Figure 4-3 A testing task – Making a decision at peer review stage



Finally the title of the testing task was kept the same as before - "*Making a decision on a submission at peer review stage"* but the task was simplified. Two persons were required to work on the task. One should act as the editor and the other should act as the reviewers. Figure 4-3 depicts that the testing task consists of three steps (subtasks):

1) assigning reviewers: the editor invites two reviewers to review a submitted article;

2) reviewing a submission: two reviewers accept the invitation, review a submitted article, post their comments and notify the editor about the completion of review via the email tool in OJS.
3) making a decision: the editor makes a decision on the submission. If the editor's decision is "Resubmit for review" or would like to get the third reviewer's opinions on the article, then the review process will go to the second round. The peer review process will be ended with a decision either 'Accepted', or 'Declined', or 'Revision required' otherwise.

This is an asynchronous collaborative team task. Editor and reviewers have to work in a team and respond to one another in time in order to complete the task in a predefined timeframe. Any late responses and mistakes made by any persons would cause either running out the time or stopping the process. The task script is presented on Appendix 10.

4.3 OJS set up and Trial testing

4.3.1 OJS set up

OJS system for this case study was set up by a lecturer who is a Linux expert in SCMS at AUT. The journal website went live and could be accessed externally at the end of S1 2009. Configuration was kept as minimum. At the front end of the online system, the templates and tools, the colour theme, text and layout styles were adopted from OJS system. Only the journal site's title and description and some users' information were input to the system. The site was called as "*AUTOJS Collaborative Teaching and Learning System*" (see Appendix 11). On the backend, a file directory system and email system were set up and configured by the lecturer and AUT's IT services. Uploaded files and emails were stored in a local Linux web server in SCMS.

4.3.2 Trial Testing

Two trial testing sessions were conducted before Pilot study. During the trial testing sessions, the proposed framework, questionnaires, the task script, and TS were checked in order to find out if they were readable and understandable, and the testing task was practicable, and what needed to be modified on them. Some changes and corrections were made on post-test questionnaire, the framework, and the task script after the two testing sessions.

Table 4-1 The progress of developing Part 1 of Post-test Questionnaire

Dimensions	1st draft			2nd draft (removed weight%)		3rd draft (removed weight%)	
	weight	Criteria (C)	# of S	Criteria (C)	# of S	Criteria (C)	# of S
Effective-ness	16%	Completeness	1	ditto	1	ditto	1
		Visibility	2	ditto	1	ditto	1
		Organisation /Design	3	ditto	2	ditto	2
		Navigability	4	ditto	2	ditto	2
		Subtotal (#C/#S)	4/10	**Subtotal (#C/#S)**	4/6	**Subtotal (#C/#S)**	4/6
Efficiency	16%	Speed	2	ditto	2	ditto	2
		Familiarity & Consistency & Standards	3	ditto	2	ditto	2
		Flexibility /Adaptability /Configurability	2	ditto	1	ditto	1
		Effort	3	ditto	3	ditto	3
		Subtotal (#C/#S)	4/10	**Subtotal (#C/#S)**	4/8	**Subtotal (#C/#S)**	4/8
Collaborativity	28%	Role Management	2	ditto	2	ditto	2
		Awareness	2	ditto	1	ditto	1
		Communication	4	ditto	1	ditto	1
		User Control /Moderator & Teacher Control	3	ditto	2	ditto	2
		File & Records Sharing /Management	4	ditto	2	ditto	2
		Process Tracking /Automated Notification	4	ditto	2	ditto	2
		Security	3	ditto	2	ditto	2
		File/Content Protection	-	added	1	ditto	1
		Subtotal (#C/#S)	7/22	**Subtotal (#C/#S)**	8/13	**Subtotal (#C/#S)**	8/13
Error Tolerance	8%	Error Rate	-	ditto	-	ditto	-
		Error Prevention	4	ditto	3	ditto	3
		Subtotal (#C/#S)	2/4	**Subtotal (#C/#S)**	2/3	**Subtotal (#C/#S)**	2/3
Accessibility	12%	Internationalisation	1	(removed)	-	(removed)	-
		Disabled User Support	1	(removed)	-	(removed)	-
		Different System Platform Support	4	(removed)	-	(removed)	-
		Support different users with different levels of IT expertise	-	added	1	ditto	1
		Subtotal (#C/#S)	3/6	**Subtotal (#C/#S)**	1/1	**Subtotal (#C/#S)**	1/1
Satisfaction	20%	Usefulness /Functionality	3	ditto	3	ditto	3
		Learnability/Predictability /Recognition /Memorability	2	ditto	2	ditto	2
		Simplicity	1	ditto	1	ditto	1
		Help/Documentation	2	ditto	3	ditto	3
		Aesthetic Design	1	ditto	1	ditto	1
		Overall	1	ditto	2	ditto	2
		Subtotal (#C/#S)	6/10	**Subtotal (#C/#S)**	6/12	**Subtotal (#C/#S)**	6/12
		Total (#C/#S)	26/60		25/43		25/43

Note: the changes on wording of the statements were not included on this table.

C: criteria; S: statements;

1st draft: was completed before trial testing; 2nd draft was completed after 1st session of trial testing; 3rd draft was completed after 2nd session of trial testing;

Post-test questionnaire consists of three parts. Table 4-1 above illustrates that the main changes were made on Part 1 of post-test questionnaire after each trial testing session. The modifications include removing duplicate, vague, and inapplicable criteria and statements, and adding some criteria and statements as needed. Here are some examples.

In the Effectiveness dimension: for "Navigation",

One old statement: "It is easy to find where I am when working on a task and find the information I needed".

The statement was changed to: "It is easy to find where I am and the information I needed when working on a task".

A statement removed: "There are various access methods in the system (a list, a drop down menu, or quick links) that enable me to begin and finish a task quickly and easily". The reason for this statement to be removed is that it was not applicable to OJS.

In the Collaborativity dimension:

for "Communication**"**, three statements below which are inapplicable to OJS were removed: "I am able to communicate with the teammates synchronously at any time"; "I am able to communicate with my teammates synchronously by sending text messages on the system"; "I am able to communicate with my teammates synchronously via audio conference provided on the system";

but a criterion - "File/Content Protection" was added into the dimension.

In the Universal Accessibility dimension:

"Disabled User Support", "Internationalisation", and "Different System Platform Support" were inapplicable criteria for OJS. They and their statements were removed from the dimension because none of disabled users and overseas users was involved in the testing and also the system was set up for windows users only. But "Support different users with different levels of IT expertise" and their statements were added into the dimension.

Some words in the open-ended questions on Part 3 were amended. For example, Qb did not specify that the question was about UE in general while OJS was not added to the end of Qc so it was unclear which system a question was referred to. There was no change made on Part 2.

Consequently, the number of criteria was changed to 25 from 26, and the number of statements decreased to 43 from 60 on Part 1. So, the draft 3 of Post-test questionnaire has 43 statements about 25 criteria in Part 1, 6 questions about importance of 24 criteria in Part 2, and 12 open-ended questions in Part 3. Part 1 has 25 criteria including 24 criteria in Part 2 and the criterion -"overall".

Table 4-2 The progress in developing the Task Script

Task type	Draft 1	Draft 2	Draft 3
1 single user task – enrolling users	One participant acts as Journal manager;	One participant acts as Editor;	(done before testing)
	5mins	ditto	
	Journal manager enrols users and assigns roles to the users	Editor enrols users and assigns roles to the users	
1 single user task - submitting an article	One participant acts as Author;	ditto	(done before testing)
	5 mins;	ditto	
	Author logon and submit an article;	ditto	
1 team task - peer review	four participants – one editor, three reviewers;	1 participant - Editor, a helper acts as three reviewers;	1 participant - Editor, the researcher acted as reviewers
	need 3 computers;	need 1 computer	ditto
	20 mins;	20 mins;	ditto
	Peer review & then Editor makes a decision on the submission according to the reviewers' recommendations. (the decision was defined on the script) The task had 2 rounds of peer review.	Peer review & then Editor makes a decision on the submission according to the reviewers' recommendations. (removed the predefined decision)	ditto

Note: 1st draft: was completed before trial testing; 2nd draft was completed after 1st session of trial testing, 3rd draft was completed after 2nd session of trial testing;

Table 4-2 lists the changes made on the task script which was created based on the task design in Section 4.2.2. The draft 3 of the task script simplified the testing task, defined one team task (a peer review) and provided the details of the steps for completing the task. The researcher/facilitator/observer should act as two or three reviewers and the participant should act as the Editor and make a decision on a submission according to reviewers' recommendations. The peer review process should be finished either at the end of the first round or the second round review within 20 minutes.

4.3.3 Summary

The Post-test questionnaire, the framework, and the task script were modified and ready for Pilot study for further checking after trial testing. The draft 3 of the questionnaire

and the draft 3 of task script not only were more workable and better fit in the testing scenario than the previous versions of the drafts, but also overcame the difficulties in:

1) scheduling 12 testing sessions which three or four participants (postgraduate students or academic staff) should meet together at the same time in a session;
2) time management i.e. it would be hard to manage time when multiple users worked on the collaborative task in different rooms or a big room within a limited time;
3) observing three or four participants by the researcher when they were working on the collaborative task via asynchronous communication in different rooms or in a big room;
4) booking three or four different testing rooms where a computer and a voice recorder need to be available in each room at a time or booking one big room where many computers and voice recorders need to be available in that room.

4.4 Participant recruitment

4.4.1 The potential participant

The potential participants were the postgraduate students and staff in SCMS at AUT except the students who were taught by supervisors and could be identified by the supervisors as defined in the Approval for AUTEC Ethical Application 09/29 (see Appendix 12). The ideal participants would be those people who had studied or were studying CC, and had got or would like to get some experience in using OJS. The reasons are

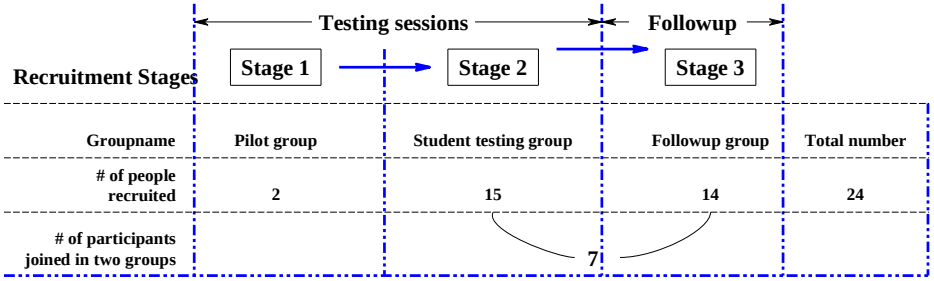
1) the system to be tested in this study is OJS. It is free but is not a university-wide system at AUT. It was used for the students who studied CC to complete their group assignments in 2009. This system was also set up as a journal publishing system and had been used by some academic staff in SCMS. So some students and staff in SCMS had got the experience in using the system.
2) one difficulty in doing a research is getting a large number of participants. How to motivate students or staff to take part in the study is the key. Not many people would like to spend their time on taking part in a research which they are not interested or motivated. The students studying CC could have more interest in participating the testing on OJS than other students who did not study the paper as taking part in the testing could help the students in CC class understand OJS and complete their group assignments on time.
3) the CC course is not taught by the supervisors.

Therefore, the participants should be recruited from the students who were studying or had studied CC, and the staff in SCMS at AUT.

4.4.2 The stages of participant recruitment

As the CC paper was offered to the students in S2 2009, the recruitment started at the end of S1 2009. Figure 4-4 shows the three stages of participant recruitment. 24 participants were recruited in this case study, 2 in the Pilot group, 15 in the student testing group, and 14 in the follow-up group (7 persons were in both testing and follow-up groups).

Figure 4-4 Three stages of participant recruitment



Stage 1*:* At the end of S1 2009, the invitation letter (shown on Appendix 13) was delivered to staff pigeon holes in SCMS. Eventually two staff joined the pilot group.

Stage 2: the CC course was taught in S2 2009. At the beginning of S2, a 30-minute training session was presented to the students in class. The researcher introduced OJS and the research, and then demonstrated how to logon and submit an article on OJS. At the end of the training session, the students were informed that OJS would be used for completing their group assignments, and encouraged to go for the testing as it would be good for them to know OJS before doing their assignments. The invitation letter, the participant information (see Appendix 14), and Consent form (see Appendix 15) were distributed to all 32 students in the class. 15 consent forms were returned to the researcher at the end of the class. Therefore, 15 students joined the student testing group.

An email with an attached spreadsheet showing timeslot options was sent to the 15 students together with an explanation how to select their preferable time slots for doing the testing in the same week when the consent forms were received. 15 students confirmed their time slots via email and then 15 testing sessions were arranged during two months. A participant's number was assigned to each person.

Stage 3: The students who studied CC were required to use OJS to work on their group assignment – collaboratively reviewing a group assignment in a team on OJS. At the beginning of the semester, a journal section for CC group assignments on OJS was created and ready for students to upload their group assignments. The students were divided into 16 teams and each team had two students. The lecturer acted as the journal manager and the editor, and enrolled all students as Authors, reviewers, and section editors. Each team (a pair) was required to submit the group assignment to OJS by one Friday, and then complete the first round of peer reviews by the following Friday.

Near to the due date of the first round of the peer review, with the permission from the CC lecturer, the researcher attended to an evening class, talked to some students individually and invited them to join the follow-up group in which the participants only needed to complete two questionnaires. 14 students including seven students who had done testing in last stage agreed to participate. 14 pre-test questionnaires and 14 post-test questionnaires were handed out to the students, and seven Consent Forms were received on that evening. An email was sent to each student in the following week to advise a participant's number and the due date of handing back the questionnaires. So, 14 students formed the follow-up group.

4.5 Pilot study

As defined in Section 3.4.2 and Section 4.2.2, a testing session consists of five steps and a testing task includes three steps. Before a testing session, an article was submitted and some users were enrolled in the OJS journal website "*AUTOJS Collaborative Teaching and Learning System"*, and reviewers' comments had already been made so reviewers just needed to copy them to OJS system in order to save time.

Two testing sessions were conducted at the stage - Pilot study. During a session, a participant acted as the editor and was given the task script, and the facilitator acted as the observer and reviewers. Both worked on the testing task according to the steps provided on the task script. In the first testing session, the participant/the editor decided to invite the third reviewer to review the article at the end of the first round of review, the task was moved on to the second round of review. But the task was not completed in the second round because the time was run out. So, the final decision on the review was not made, and the interview questions were not asked. In the second testing session, the participant made a decision and finished the task at the end of the first round review,

then completed the questionnaires and a TS, and was interviewed. The time spent on this session was about 60 minutes. The whole session was run smoothly and successfully.

A Pilot study is "*a small trial run of the main study, aims to make sure the designed plan is viable before doing the main study. Normally the study checks the procedure, interview scripts and questions, questionnaires, experiments, etc, and finds out if they are working properly and need to be modified*" (Kurfess, 2005). The pilot study had tested the task script, the TS and questionnaires which were modified after the trial testing. The feedbacks from the Pilot study were positive and no suggestions on modifying them were received. So the task script, the TS, and the two questionnaires were finalised and ready for the full-scale study.

4.6 Full-scale study

The full-scale study consisted of two groups: student testing group and follow-up group. A participant in the student testing group completed a testing task based on the task script, filled in the two questionnaires (pre-test and post-test questionnaires) and a TS, and answered the questions asked during a one to one interview. His/her feedbacks about the framework and OJS CL usability were collected. None of the participants in the student testing group had used OJS before, so their feedbacks would be just based on a short time (about 20 minutes) experience with OJS.

As explained in Section 4.4.2, the follow-up group were recruited after the students in the CC class had started working on their group assignments and peer reviewing assignments on OJS. A participant was required to complete the two questionnaires (pre-test and post-test questionnaires) only.

4.6.1 Student testing group (ST group)

A dedicated testing room and time schedules for using the room were arranged and a laptop was borrowed from SCMS at AUT. In each testing session, two persons were involved in the team task. One was a participant who acted as the Editor. The other was the facilitator who acted as the observer and the reviewers.

The 15 participants who signed the consent form were asked to come to the testing room, were given the task script which shows the steps of completing the task, and were

required to complete the testing task at the time slots they selected. However, two students did not turn up at the time slots they chose without giving any notices. Three students including one student who withdrew the paper study sent emails about cancelling their appointments and did not take part in the testing. So, 10 testing sessions were conducted over a five-week period. The shortest time spent on the testing task was 12mins while the longest time was 33mins. More details on time spending on each testing session are shown in Appendix 18. During a testing session, a participant completed a testing task, filled in the two questionnaires and a TS, and was interviewed.

4.6.2 Follow-up group (FU group)

Fourteen students joined the FU group. They were not required to do the testing task but needed to complete the questionnaires because they had worked on peer review of their group assignments and had got some experience in using OJS. Pre-test questionnaire and Post-test questionnaire were handed out to the 14 participants on an evening class. It was supposed to get the questionnaires back in two weeks after they were handed out. However, collecting the questionnaires ended up as a long process. It took one month to collect the questionnaires from the 13 participants. Several students did not hand in the questionnaires until the fourth reminder email was sent out. One student never returned the questionnaires. Several participants did not answer all the questions.

4.6.3 Summary

All participants in full-scale study were recruited from the CC class. 10 testing sessions were conducted and each session was participated by a participant in the ST group. The FU group had 13 participants and each participant only completed Pre-test questionnaire and Post-test questionnaire. The ST group's working environment was different from the FU's. The participants in ST group had a preset working environment and accessed OJS and completed a predefined task within a limited time whereas the participants in the FU group accessed OJS either at home or the school at their own time and spent more than a month on working on OJS to finish peer review and complete their group assignments.

4.7 Data collection

4.7.1 Data collection from the testing sessions

The pilot study group and the ST group took part in OJS usability testing. User observation was undertaken during the testing sessions. At the beginning of a testing session, the participant was told that he/she should apply Think Aloud, i.e. should keep

talking about what he/she was working on and what thoughts, questions, and comments came up to his/her mind during the testing. The facilitator/observer watched what he/she was doing and filled in an ODCS to record the time spent on the task, mistakes, comments and questions heard from the participant, and so on (ODCS is shown on Appendix 17). Meantime, the observer used Camtasia studio to capture screen and voice, and turned on a digital voice recorder to record the participant's talk. A TS, a Pre-test questionnaire, and a Post-test questionnaire were completed by a participant and handed back to the facilitator at the end of a testing session. An ODCS was completed by the facilitator/observer.

Therefore the collected raw data consisted of three types of data - digital video, digital audio, and handwritten data. Each participant was given a participant's number and he/she was required to write down the number on his/her TS and questionnaires. The participant's number and the date of doing a testing session were used to name the data files.

4.7.2 Data collected from Follow-up group

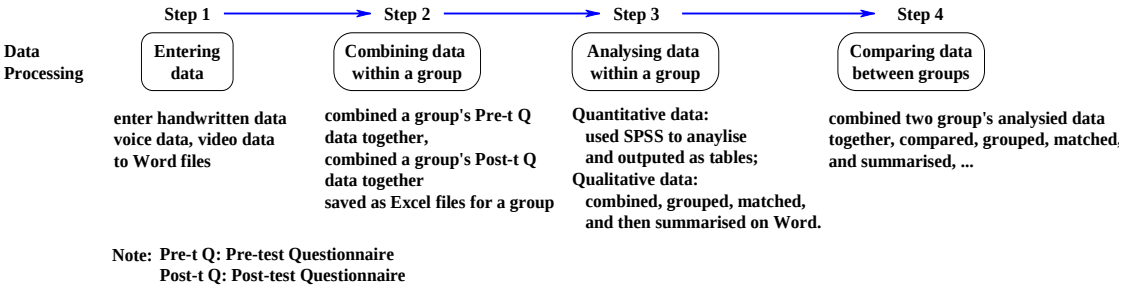
Thirteen participants in the follow-up group complete the two questionnaires – Pre-test and Post-test questionnaires. Each participant was given a participant's number and was required to write down the number on each questionnaire so the number could be used to name the data files. 13 Pre-test and Post-test questionnaires were collected. Therefore the collected raw data has one type of data, i.e. handwritten data.

4.8 Data processing

Any research needs multiple points of view and triangulation in order to get rich and converged data and avoid subjective data (Hollingsed & Novick, 2007; Scholtz, 2004). As explained in last Section 4.7, the source data collected from the pilot study and the full scale study consists of three types – handwritten data, digital voice data, and digital video data. Some data is qualitative and some data is quantitative.

Yin (1994, 2003a) suggested that "*logic models, time-series analysis, pattern-matching, explanation-building, and address rival explanations*" are the common techniques of data analysis for case study. This research had applied some of these techniques and used MS office and SPSS to tabulate and analyse the data.

Figure 4-5 Four Steps of data processing



There are four steps of data processing as shown on Figure 4-5.

Step 1: entering the raw data collected from each participant to Word files.

Handwritten responses were entered in Word files. A digital voice file for an interview was transcribed into a Word file. A digital video file (.wmv file) and a audio file, that captured the starting time and the end time of a testing session, errors made by a participant, the questions that he/she asked during his/her testing session, and so on, were transcribed and input to a Word file. The data could be supplementary data for a TS and an ODCS. So, the data was then added into a file that contains the data gathered either from a TS or an ODCS as needed. For each participant in the ST group and the pilot study group, five files were created that contained the data from a TS, an ODCS, a Pre-test questionnaire, a Post-test questionnaire, and an interview respectively. For each participant in the follow-up group, two files were created for Pre-test questionnaire and Post-test questionnaire separately.

Step 2: combining source data together within a group.

Data samples were divided into two groups. One group was called as "Testing group (TG)" and consisted of the ST group and the pilot study group. The other was called "Follow-up group (FU)".

The reasons of putting the ST group and the pilot study group together were:

- the participants from both groups completed exactly the same tasks: usability testing, the questionnaires and sheets, and an interview during a testing session;
- the size of either groups was small. Combination of both groups increased the size of the data sample, from 10 to 12. This would be beneficial to the data analysis.

Within TG, 12 participants' files from one source, e.g. Pre-test questionnaire, were combined into one file while in the follow-up group 13 participants' responses to Pre-test questionnaire were put together. The raw data from TG's TSs and ODCS were combined together and saved in a Word file while the data from interviews for TG were

input into a Word file. For each group, the data about open-ended questions in Part 3 of Post-test questionnaire was combined to a Word file. The data collected from Part 1 and Part 2 of Post-test questionnaire were stored in two different Excel files for each group.

Step 3: processing raw data within a group.
Most quantitative raw data was collected from Pre-test questionnaire, and Part 1 and Part 2 of Post-test questionnaire. Some data from Pre-test questionnaire was calculated in MS Excel and generated the outputs such as percentages, mean, max, min, etc.

Post-test questionnaire includes two parts of quantitative data. Part 1 uses a five-point Likert scale ranging from "strongly disagree (1)" to "strongly agree (5)" and Part 2 uses four-point Likert scale ranging from "not important (1)" to "very important (4)". The data of Part 1 and Part 2 from the two groups were imported into SPSS separately. Responses like "Not Applicable (N/A)" or "No response (N/R)" were treated as missing data and were excluded from the calculations in SPSS. This had affected the size of the data sample, e.g., if for one statement, two participants selected N/A, then the number of the total participants responded to this statement would be 12-2=10 in TG or 13-2=11 in FU. "Frequencies, Custom Tables, and Report summaries in row" were calculated and "Min, Max, Mean, % of participants selected a scale" were outputted as findings which will be presented in chapter 6.

The qualitative raw data was collected from the TS, ODCS, the open-ended questions in Part 3 of Post-test questionnaire, and the interviews. Data had been entered into Word files at Step 2. Firstly, the data on a file were sorted and grouped, e.g. data about OJS was kept together. Positive responses were separated from negative responses. Secondly, numbers or percentages of participants who gave the same or similar comments were counted and entered in tables. Thirdly, the responses to a question/some questions were synthesised, categorised, and summarised. Finally four files were created. Two files contained the data on the questions in Part 3 of Post-test questionnaire for TG and FU respectively. The other two contained the data from the TSs and ODCSs, and interviews separately for TG.

Step 4: combining the TG and FU's data together for data analysis and comparison. The data sets from the two groups were put together in tables, where they could be

compared, grouped, and summarised to produce findings which will be explained in Chapter 6.

Table 4-3 summarises what raw data was collected and what outputs were produced from the data processing.

Table 4-3 Types of the raw data collected and the outputs of data processing

Raw data collected					**Data output**	
Types of raw data	**Groups**	TG		FU	TG	FU
		Pilot study	ST group	Follow-up group		
	# of a source collected	#	#	#		
	Names of data sources					
Handwritten data	TS	2	10	-	qualitative data in a Word file	-
	ODCS	2	10	-		
	Pre-test questionnaire	2	10	13	qualitative & quantitative data in an Excel file for a group	
	Part 1 of Post-test questionnaire	2	10	13	quantitative data in a SPSS output file was transferred to a excel file for a group	
	Part 2 of Post-test questionnaire	2	10	13	quantitative data in a SPSS output file was transferred to a excel file for a group	
	Part 3 of Post-test questionnaire	2	10	13	Qualitative data in a Word file for a group	
Video data	Records of think aloud	2	10	-	Qualitative data was entered into the Word file which contained the data from TS & ODCS	-
Voice data	Records of Think aloud	2	10	-		-
Voice data	Records of an interview	2	10	-	qualitative data was entered to a Word file	
Handwritten data	interview	2	10	-		

Note: **TS**: Task sheet; **ODCS**: Observer data collection sheet;

4.9 Summary

OJS is a free online journal publishing system and has been used by many universities around the world. It was set up as a CSCL system in SCMS for the postgraduate students who studied CC to collaboratively work on their group assignments. In this case study, OJS was employed as a CSCL system to be evaluated by using the framework which was proposed at stage 2 of this study and presented in Section 2.6.

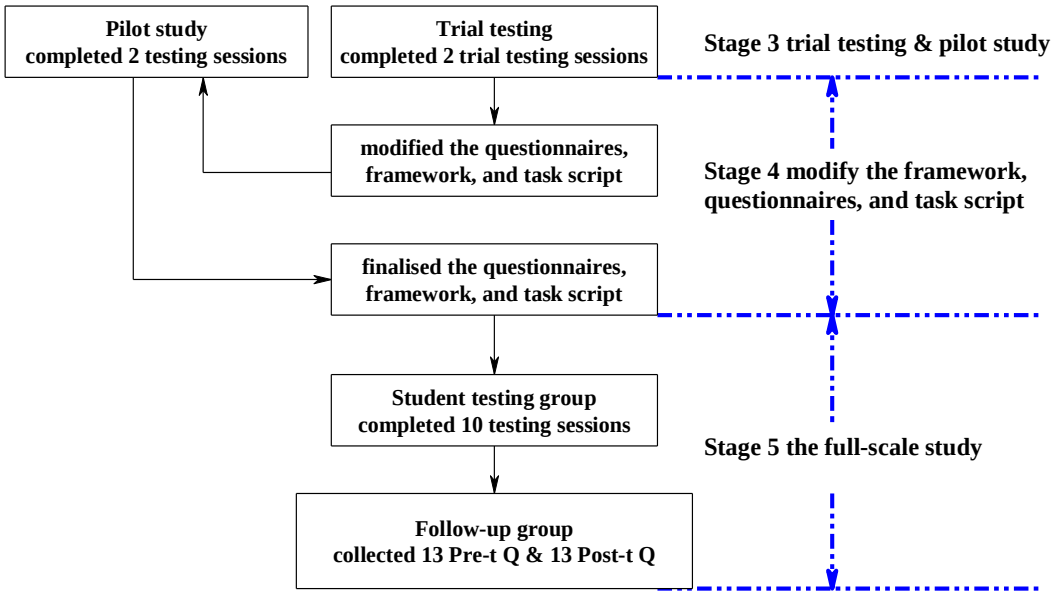
A testing session consisted of five steps: introduction, completing Pre-test questionnaire, working on the testing task, completing Post-test questionnaire, and an interview. The testing task is a collaborative team task - peer review and making a decision on a submission. A participant was the Editor, and the facilitator acted as the observer and the reviewers. The Editor needed to make a decision based on the comments and the

recommendations provided by the reviewers. A participant was required to use Think Aloud Protocol during a testing session. The talk and screen activities were recorded and captured as digital audio and video data. The questionnaires and TS and the ODCS were collected at the end of a session.

In Trial testing, the framework, the task script, and the questionnaires were tested and then were corrected and modified. Then they were tested again and finalised in the pilot study, and were used in the full-scale study which involved a ST group and a FU group. Twenty four persons were recruited but 18 persons participated the study. Twelve testing sessions (including two trial testing sessions) were successfully conducted although some issues and problems occurred during the period (this will be discussed in Section 7.5).

It took four and a half months to complete the OJS usability testing and data collection, and to finish the three stages - stage 3, 4, and 5 defined in Section 3.3.1 (see Figure 4-6). Three types of source data were collected, including 12 TS and 12 ODCS, 12 files of containing digital audio and video data separately, and 25 pre-test questionnaires and 25 post-test questionnaires (see Figure 4-7 next page). There are four steps of data processing - from raw data to Word and Excel files and SPSS files: entering the raw data collected from each participant to Word files, combining source data within a group, processing raw data within a group, and combining the TG and FU's data together for data analysis and comparison.

Figure 4-6 The progress of OJS Usability Testing and Data Collection



Note: **Pre-t Q: Pre-test Questionnaire**
Post-t Q: Post-test Questionnaire

Figure 4-7 Summary of Data collection

	Testing sessions		Followup	
Group name	**Pilot group**	**Student testing group**	**Followup group**	**Total number**
# of people recruited	2	15	14	24
# of people participated	2	10	13	18
# of participants joined in two groups		7		7
Tasks/files were completed				
1) a testing task on OJS	Yes	Yes	Not	12
2) Task Sheet	Yes	Yes	Not	12
3) Interview	Yes	Yes	Not	12
4) Observer Data Collection Sheet	Yes	Yes	Not	12
5) Pre-t Q	Yes	Yes	Yes	25
6) Post-t Q	Yes	Yes	Yes	25

Note: **Pre-t Q: Pre-test Questionnaire**
Post-t Q: Post-test Questionnaire

Chapter 5

5. Six dimensions internal consistency analysis

This chapter explains the theory of Internal Reliability Analysis and Inter-Rater Reliability analysis, and then examines if the rankings of agreements on the statements of the criteria in each dimension are internal consistent, if the rankings of importance of the criteria in each dimension to system usability is consistent, and also if the consistency of participants' responses to the questions is acceptable. Then it discusses the outcomes of the analysis and addresses a concern about the reliability analysis in this study. The following six sections are presented in this chapter.

5.1 Reliability analysis in theory

5.2 Internal Reliability (IR) analysis in this study

5.3 Inter-Rater Reliability (IRR) analysis in this study

5.4 Discussion

5.5 The concern on the reliability analysis in this study

5.6 Summary

5.1 Reliability analysis in theory

5.1.1 Reliability

"*Whenever you use humans as a part of your measurement procedure, you have to worry about whether the results you get are reliable or consistent"* (Trochim, 2006). Reliability *"means that a scale should consistently reflect the construct it is measuring*" (Field, 2005) and "*has to do with the quality of measurement"* (Trochim, 2006). Reliability analysis *"can be used to measure the consistency of questionnaire*" (Field, 2005). There are "*four general classes, each of which estimates reliability in a different way".* They are Inter-Rater or Inter-Observer Reliability ("*assessing the degree to which different raters/observers give consistent estimates of the same phenomenon")*, Test-Retest Reliability ("*measuring from one time to another"*), Parallel-Forms Reliability **("***assessing two tests constructed in the same way from the same content domain"*), and Internal Consistency Reliability *("assessing the consistency of results across items within a test"*) (Trochim, 2006). This thesis will focus on Internal Reliability analysis and Inter-Rater Reliability analysis.

5.1.2 Internal Reliability and Cronbach's Alpha (α)

Bryman and Cramer (2009) advised that internal reliability is particularly important in connection with multiple-item scales as it checks whether the items that make up the

scale are internally consistent. The most widely used measure for estimating the internal consistency of multi-item tests is Cronbach's alpha (Field, 2005; McGraw & Wong, 1996; Santos, 1999).

Cronbach's Alpha is *"a test reliability technique that requires only a single test administration to provide a unique estimate of the reliability for a given test*" (Gliem & Gliem, 2003) and measures "*how closely related a set of items are as a group"* (UCLA*), and* "*determines the internal consistency"* (Santos, 1999). If a questionnaire has subscales, Alpha (*α*) "*should be applied separately to these subscales*" (Field, 2005). Bryman and Cramer (2009) also recommended that "*when a concept and its associated measure are deemed to comprise underlying dimensions, it is normal to calculate reliability estimates for each of the constituent dimensions rather than for the measure as a whole"*.

Three important values generated from Internal Reliability analysis in SPSS are:

1) *Cronbach's alpha reliability coefficient*: the Alpha coefficient *"can take values between negative infinity and 1 (although only positive values make sense)*" (*Cronbach's alpha*) but *"normally ranges between 0 and 1"* (Gliem & Gliem, 2003) and is shown in a Reliability Statistics table. Field (2005) explained that Cronbach's Alpha (α) is "*the overall reliability of the scale"* and suggested that one should look for "*the values in the magnitude of .7 to .8 (or thereabouts).* The rule of thumb is *">.9* (reliability) *Excellent, >.8 Good, >.7 Acceptable, >.6 Questionable, >.5 Poor, and <.5 Unacceptable"* (Gliem & Gliem, 2003; Lani, 2009). In most social science research situations*, "a reliability coefficient of .70 or higher is considered 'acceptable'"*(UCLA) but *"lower thresholds are sometimes used in the literature"*(Santos, 1999). "*A very high alpha (e.g. >.90) probably means that the items are repetitious or that you have more items in the scale than are really necessary for a reliable measure of the concept"* (Leech, Barrett, & Morgan, 2005).
2) *Alpha if Item Deleted*: are "*the values of the overall* ***α*** *if that item is not included in the calculation. … If the deletion of an item increases Cronbach's* ***α*** *then is means that the deletion of that item improves reliability. Therefore, any items that result in substantially greater values of* ***α*** *than the overall* ***α****, may need to be deleted from the scale to improve its reliability"* (Field, 2005).

3) *Corrected Item-Total Correlations***:** the values are *"the correlations between each item and the total score from the questionnaire. In a reliable scale all items should correlate with the total. ... if any of these values are less than about* ***.3*** *(depends slightly on the sample size - with bigger samples smaller correlation coefficients are acceptable) then … it means that a particular item does not correlate very well with the scale overall. Items with low correlations may have to be dropped*" (Field, 2006). Leech, Barrett, and Morgan (2005) also suggested that if an item's value of Corrected Item-Total Correlation is 0.40 or above, then the correlation is moderately high or high. But if an item's value is negative or too low (less than 0.30), then one should examine the item for wording problems and conceptual fit, and if it should be modified or deleted.

So, if an Alpha shows poor reliability, *"then individual items within the scale must be re-examined and modified or completely changed as needed"* (Santos, 1999) and "*a strategy for dealing with this eventuality is to drop one item or more from the scale in order to establish whether reliability can be boosted*" (Bryman & Cramer, 2009).

However, Cronbach's alpha "*is not a panacea, just like any other statistic, it must be used with caution*" (Cortina, 1993). When checking data reliability, one should consider the followings:

1) "*any judgment of adequacy, even in research, needs to consider context"* (Hollenbeck, 1991, as cited in Cortina, 1993). Cronbach's alpha reliability coefficient should be .8 or above normally when the internal reliability is good, but lower than .7 could be acceptable in some social contexts.
2) The number of items in the scale affects the overall Alpha value. "*A high value for Cronbach's alpha indicates good internal consistency of the items in the scale, it does not mean that the scale is unidimensional*" (Gliem & Gliem, 2003). "*It is possible to get a large value of* ***α*** *because you have a lot of items on the scale, not because your scale is reliable*" (Field, 2005). Cortina (1993) also advised that "*alpha is very much a function of the number of items in a scale, it must be interpreted with number of items in mind"* because "*alpha can be high in spite of low item intercorrelations and multidimensionality*".
3) the number of the items in a scale should not be more than 19. Gliem and Gliem (2003) advised that "*while increasing the value of alpha is partially dependent upon the number of items in the scale, it should be noted that this has diminishing return"*. Previous research on manipulating numbers of items between 10 and 75

proved that "*the relationship between number of items and alpha is curvilinear (Komorita & Graham, 1965) and begins to level off before the number of items reaches 19*" (Cortina, 1993).

4) Alpha may be negative due to three possible reasons. The first reason can be coding or data entry errors, i.e. "*the scale consists of some items that are worded in opposite directions to alleviate response biases, and the researcher has forgotten to appropriately recode the reverse scored items, resulting in negative covariances where the actual covariances of interest are positive"*(Nichols, 1999). The second one is most likely with small sample sizes and small numbers of items. The third reason is that the items are not measuring the same thing i.e. they represent more than one dimension of meaning, and these dimensions are negatively correlated (Garson, 2010; Nichols, 1999).

5.1.3 Inter-Rater Reliability (IRR) and Intraclass correlation (ICC)

"*Any qualitative assessment using two or more researchers must establish interrater reliability to ensure that the results generated will be useful*" (Shuttleworth, 2009). Inter-Rater Reliability refers to different responses to the same questions among the raters (Intraclass correlation). It is used to "*address whether judges rank order targets in a manner that is relatively consistent with other judges"* (LeBreton & Senter, 2008). "*In the data setup, judges are the columns and judgees are the rows*" and "*estimation is based on the correlation of scores between/among two or more raters who rate the same item, scale, or instrument"*(Garson, 2010).

"*Reliability emphasizes the relative consistency or the rank order similarity between judges and is typically indexed via some form of a correlation coefficient*" (LeBreton & Senter, 2008). Intraclass correlation (ICC) can be calculated in SPSS for measuring *"inter-rater reliability for two or more raters when data may be considered interval level. It may also be used to assess test-retest reliability. ICC may be conceptualized as the ratio of between-groups variance to total variance" and "is preferred over Pearson's only when sample size is small (< 15)*"(Garson, 2010). There are six forms of ICC, that is, One-way random effects model, Two-way random effects model and Two-way mixed model; each model has two types: absolute agreement and consistency (McGraw & Wong, 1996; Nichols, 1998).

"*Each model has two versions of the intraclass correlation coefficient*. *Single measure reliability is individual ratings constitute the unit of analysis, is the one of usual interest,*

used to assess if the ratings of one judge are apt to be the same as for another judge" while *"Average measure reliability is the mean of all ratings is the unit of analysis. That is, average measure reliability gives the reliability of the mean of the ratings of all raters"*. Average measure reliability is *"close to Cronbach's alpha …* (and) *requires a reasonable number of judges to form a stable average"* (Garson, 2010).

A two-way mixed consistency model with average measure reliability was used in this study. The reasons are:

1) Two-Way Mixed is used when people effects are random and the item effects are fixed. Nichols (1998) and Garson (2010) suggested that if the k raters are a random sample from a larger population and all judges rate all targets, then two-way mixed is used. *"In the mixed model, inferences are confined to the particular set of raters used in the measurement process*", and measures of consistency should be used when *"systematic variability is considered irrelevant"*(Nichols, 1998).
2) In this study, a participant was randomly recruited from 32 students in CC class and he/she completed the Post-test questionnaire independently. One's responses are not relevant to others'. The purpose of this IRR analysis is to find out if the participants' responses are consistent.

Normally a value of ICC is from 0 to 1. That is, ICC *"will be high when any given row tends to have the same score across the columns (which are the raters)"* and *"there is perfect inter-rater reliability when it is 1.0,* (i.e.) *there is no variance within targets (for any target, all raters give the same ratings"* (Garson, 2010). However, "*it is possible for empirical estimates to be negative*" (Nichols, 1998). "*ICC < 0.4 represents poor reliability, 0.4 ≤ ICC ≤ 0.75 represents fair to good reliability, and ICC > 0.75 represents excellent reliability"*(Melzer, Shtilman, Rosenblatt, & Oddsson, 2007)..

5.2 Internal Reliability (IR) Analysis in this study

Data was divided into two groups, i.e. TG and FU as described in Section 4.8 Step 2. The statements and criteria are grouped into six dimensions/subscales – Effectiveness, Efficiency, Collaborativity, Error Tolerance, Universal Accessibility, and Satisfaction. Each group's data collected from Part 1 and Part 2 of Post-test questionnaire was analysed for each dimension separately in SPSS as per Bryman and Cramer (2009), and Field's suggestions (2005) and the outputs will be presented in Sections 5.2.1and 5.2.2.

Universal Accessibility dimension has one criterion. SPSS cannot process reliability analysis when there is only one item in a scale. As explained in Section 4.8 Step 3, if the responses were "Not Applicable (N/A)" or "No Response (N/R)", they were treated as 'missing data' and were excluded from the data process in SPSS. All the outputs of reliability analysis about Part 1 and Part 2 are listed in Appendix 19 and Appendix 20.

Here are some notes helping with interpreting the tables presented in this section:

- Number of Items: the number of the statements or criteria in a dimension.
- Valid Number of Cases: the number of participants whose responses for a dimension were processed.
- Cronbach's Alpha (**α**): the value for overall reliability of the scale. This thesis interprets a value based on the rule: "*9 (reliability) Excellent, >.8 Good, >.7 Acceptable, >.6 Questionable, >.5 Poor, and <.5 Unacceptable*" (Gliem & Gliem, 2003; Lani, 2009); and only looks for the values over .7 (reliability acceptable or better). (see Section 5.1.2 for more details).
- Alpha if Item Deleted: a value for an item, is the value of the overall **α** if that item is not included in the calculation. "*any items that result in substantially greater values of **α** than the overall α, may need to be deleted from the scale to improve its reliability*" (Field, 2005). This thesis only looks for the values over .7 (reliability acceptable or better). (see Section 5.1.2 for more details)
- Corrected Item-Total Correlations: a value for an item, is the value of the correlations between the item and the total score from the dimension (scale) that the item is included. If a value is less than about .3 and the item's Alpha if Item Deleted is over .7, then the item may have to be dropped from its dimension in order to improve the internal reliability of the dimension. (see Section 5.1.2 for more details)

5.2.1 IR Analysis – Part 1 of Post-test Questionnaire

In Part 1, a participant was required to select one of five Likert scales that best reflected to what degree he/she agreed to a statement on a criterion measuring the features of OJS CL usability. Table 5-1 lists the Cronbach's Alpha values for each dimension. The outcomes show that Effectiveness (six statements) and Efficiency (eight statements) dimensions have either acceptable or good internal consistency as their Alpha values are over .7. Error Tolerance has three statements, the TG's dataset shows it has acceptable internal consistency but the FU's dataset shows the consistency is excellent. The

Collaborativity dimension has 13 statements. Its internal consistency is questionable (shown in TG's dataset) or unacceptable (shown in FU's dataset). The Satisfaction dimension has 12 statements, and had three cases from TG and seven cases from FU calculated in SPSS. TG's Alpha value for this dimension indicates that the consistency is unacceptable whereas FU's shows the consistency is excellent.

Table 5-1 IR Analysis for the six dimensions - Part 1

	Measures	Cronbach's Alpha		Number of Items		Valid Number of Cases	
	Groups	TG	FU	TG	FU	TG	FU
Dimensions	Effectiveness	.855	.783	6	6	11	10
	Efficiency	.859	.889	8	8	7	9
	Collaborativity	.658	.409	12	10	3	3
	Error Tolerance	.766	.938	3	3	6	9
	Universal Accessibility	-	-	-	-	-	-
	Satisfaction	.476	.955	10	12	3	7

Note: number of Items: number of statements.

Table 5-2 The Items may need to be removed from the dimensions – Part 1

	Groups	TG		FU	
Statement#/Criterion	Measures	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
	Q1.1.1. Completeness	.339	.902	.237	.820
	Q2.1.2. Speed	-.121	.902	.169	.909
	Q3.4.1. User Control/Moderator & Teacher control	-.500	.720	-.115	.474
	Q3.6.1. Process Tracking /Automated Notification	-.292	.738	-	-
	Q3.6.2. Process Tracking /Automated Notification	-.292	.738	.693	.173

Table 5-2 illustrates that both groups' alpha values if Item deleted for Q2.1.2 in the Efficiency dimension are higher than their overall Alpha values (0.859 for TG) and (0.889 for FU). Both groups' values of Corrected Item-Total Correlation for Q2.1.2 are lower than 0.3. So, Q2.1.2 may need to be removed from both groups' Efficiency dimension in order to increase the dimension's internal reliability. Similarly, in TG, Q3.4.1, Q3.6.1 and Q3.6.2 from the Collaborativity dimension may need to be removed from their dimensions, while in FU, Q1.1.1 may need to be removed from the Effectiveness dimension.

However only three cases' data (i.e. three participants' responses) from each group for Collaborativity dimension, and three cases' data from TG for Satisfaction dimension were calculated in SPSS because the rest cases for the dimensions had many "N/A" or

"N/R" and were treated as missing data and excluded from the data process. The numbers of the cases for the two dimensions were small.

5.2.2 IR Analysis – Part 2 of Post-test Questionnaire

In Part 2, a participant was asked to choose one of four Likert scale to indicate how important they believed each criterion to CSCL system usability was.

Table 5-3 presents the Cronbach's Alpha values for each dimension. Effectiveness and Efficiency dimensions have four criteria each. Their overall Alpha values are negative or very small (less than 0.4). Therefore their internal consistencies could be unacceptable. The Error tolerance dimension has two criteria only. Its internal reliability is questionable for TG but is good for FU. The Collaborativity dimension has eight criteria, TG's internal reliability is good but FU's is questionable. The Satisfaction dimension has five criteria and their internal consistency for each group is acceptable.

Table 5-3 IR analysis for the six dimensions – Part 2

	Measures	Cronbach's Alpha		Number of Items		Valid Number of Cases	
	Groups	TG	FU	TG	FU	TG	FU
Dimensions	Effectiveness	.207	-.865	4	4	12	13
	Efficiency	-.012	.303	4	4	12	13
	Collaborativity	.803	.628	8	8	10	11
	Error Tolerance	.667	.807	2	2	11	10
	Universal Accessibility	-	-	-	-	-	-
	Satisfaction	.755	.748	5	5	10	13

Note: number of items: number of criteria.

Table 5-4 The Items may need to be removed from the dimensions – Part 2

	Groups	TG		FU	
	Measures	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Question#/ Criterion	Q3a3 Communication	.000	.846	.183	.631
	Q3a8 Security	.633	.769	-.183	.710
	Q6a4 /Help/Documentation	.124	.877	.672	.642

Table 5-4 illustrates that removing Q3a3 (*Communication*) from TG's Collaborativity dimension, and removing Q3a8 (*Security*) from FU's Collaborativity dimension would make the rankings of the importance of the criteria in the dimension more consistent. Dropping Q6a4 (*Help/Documentation*) off from TG's Satisfaction dimension would improve the internal consistency of the rankings of the importance of the criteria because these the criteria Cronbach's Alpha values if Item Deleted are higher than their

dimensions' overall Alpha's values and over .7, and their Corrected item-total Correlation values are small than .3.

5.3 Inter-Rater Reliability (IRR) Analysis in this study

This thesis adopted a two-way mixed consistency model with average measure reliability to analysis IRR in SPSS as explained in Section 5.1.3. The following two Sections 5.3.1and 5.3.2 will illustrate the outputs of IRR analysis for Part 1 and Part 2 of Post-test questionnaire in SPSS. (The full data outputs are shown on Appendix 21 and Appendix 22). Participants are the columns/variables/Item and the criteria are the rows/cases. Similarly to IR analysis in Section 5.2, the responses such as "N/A" and "N/R" are treated as missing data, so these responses were excluded from the data analysis process and resulted in some items having zero variance in SPSS. The Error Tolerance and Universal Accessibility dimensions had too few cases/criteria (e.g. case N = 0), and execution could not be completed. Therefore, no output was generated for these two dimensions.

Some notes for the tables are listed below for helping with interpreting the tables presented in this section:

- Interpreting ICC value is based on the rule: "*ICC < 0.4 represents poor reliability, 0.4 ≤ ICC ≤ 0.75 represents fair to good reliability, and ICC > 0.75 represents excellent reliability"* (Melzer, et al., 2007).

Table - "Intraclass Correlation Coefficient" - Average Measures:

- [a] means "Type C intraclass correlation coefficients using a consistency definition - the between-measure variance is excluded from the denominator variance";
- [b] means "this estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise".

5.3.1 IRR Analysis – Part 1 of Post-test questionnaire

Table 5-5 below shows that the dimension that has the highest number of criteria excluded from the data analysis process is Collaborativity, that is, TG has nine excluded cases and FU has eight. The second one is Satisfaction. Both groups' ICC coefficient values show that the Collaborativity dimension's IRR are excellent and the Effectiveness dimension's are fair or near good but the Satisfaction dimension's are poor. For Efficiency dimension, TG's IRR is excellent and FU's IRR is fair.

Table 5-5 IRR Analysis output – Part 1

Case Processing Summary - Cases N

	Groups	TG			FU		
	Cases N	Valid	Excluded	Total	Valid	Excluded	Total
Dimensions	Effectiveness	4	2	6	4	2	6
	Efficiency	6	2	8	7	1	8
	Collaborativity	4	9	13	5	8	13
	Error Tolerance	-	-	-	-	-	-
	Universal Accessibility	-	-	-	-	-	-
	Satisfaction	7	5	12	6	6	12

Note: cases N: number of statements.

Intraclass Correlation Coefficient - Average Measures

	Groups	TG	FU
	Measures	Intraclass Correlation[a]	Intraclass Correlation[a]
Dimensions	Effectiveness	.536[b]	.604[b]
	Efficiency	.861[b]	.491[b]
	Collaborativity	.882[b]	.848[b]
	Error Tolerance	-	-
	Universal Accessibility	-	-
	Satisfaction	.055[b]	-.086[b]

5.3.2 IRR Analysis – Part 2 of Post-test questionnaire

Table 5-6 IRR Analysis output – Part 2

Case Processing Summary - Cases N

	Groups	TG			FU		
	Cases N	Valid	Excluded	Total	Valid	Excluded	Total
Dimensions	Effectiveness	4	0	4	4	0	4
	Efficiency	4	0	4	4	0	4
	Collaborativity	6	2	8	5	3	8
	Error Tolerance	-	-	-	-	-	-
	Universal Accessibility	-	-	-	-	-	-
	Satisfaction	4	1	5	5	0	5

Note: cases N: number of criteria.

Intraclass Correlation Coefficient - Average Measures

	Groups	TG	FU
	Measures	Intraclass Correlation[a]	Intraclass Correlation[a]
Dimensions	Effectiveness	-1.493[b]	.542[b]
	Efficiency	-.545[b]	.564[b]
	Collaborativity	.533[b]	.319[b]
	Error Tolerance	-	-
	Universal Accessibility	-	-
	Satisfaction	.854[b]	.227[b]

Table 5-6 above shows that the dimension that has the highest number of criteria excluded from the data process is Collaborativity, that is, TG has two excluded cases and FU has three. TG's ICC values are different from FU's. For Effectiveness and

Efficiency dimensions, FU's IRR values are fair but TG's are poor. For Satisfaction and Collaborativity, TG's IRR values are excellent and fair respectively but FU's are poor.

5.4 Discussion

Table 5-7 on pg 82 shows Internal Reliability (IR) status and Table 5-8 on pg 83 illustrates Inter-Rater Reliability (IRR) status. Both tables list the items which may need to be removed from the dimensions that they belong to.

About OJS CL usability, the main outcomes from the IR and IRR analysis based on the data collected from Part 1 of Post-test questionnaire are:

IR:

The agreements on statements on in Effectiveness, Efficiency, and Error Tolerance dimensions were internally consistent. Q2.1.2 statement (i.e. "*The System speed is fast enough"*) may need to be removed in order to increase the reliability of the Efficiency dimension. In FU, the agreements on statements in the Satisfaction dimension were excellently consistent but in TG the agreements on statements in the same dimension were inconsistent. This implies that TG's participants thought that either the features of OJS usability reflected by the criteria and the statements were not available or the criteria and the statements did not measure the same things in the Satisfaction dimension. So the criteria and the statements in the Satisfaction dimension might need to be retested or revised or restructured. Similarly, both groups' analysed data shows that the agreements on statements in the Collaborativity dimension were internally inconsistent without considering the small number of cases that were processed. The criteria and the statements in the dimension might need to be retested or revised or restructured.

According to a group IR analysis, Q3.4.1 (*Being a moderator, I am able to give online instructions, and monitor teamwork on the system*), Q3.6.1 (*able to send a notification to the team after have completed a task*), and Q3.6.2 (*able to find out the status of a task/teamwork*) might need to be removed from the Collaborativity dimension while Q1.1.1 (*complete a task on this system within a proper time frame*) might need to be removed from the Effectiveness dimension and Q2.1.2 (*The System speed is fast enough*) might need to be removed from Efficiency dimension in order to increase the dimensions' internal consistency.

Because of missing data, the statements were dropped off by SPSS are: in the Collaborativity dimension: Q3.1.1 (*easy to add /assign users, and manage user roles/accounts on the system*), Q3.6.1 (*able to send a notification to the team after completion of a task*), and Q3.7.1 (*file sharing and protection*) (based on FU's data); Q3.4.2 (*able to manage my files/notes and the shared files/notes*) (based on TG's data). In the satisfaction dimension: Q6.2.2 (*Tasks can be performed in a straight-forward manner*) and Q6.4.2 (*easy to access help documents)* (based on TG's data) (shown Table 5-7 pg82). This implies that the participants in a group might have considered the items were not applicable to OJS. However help/documents (Q6.4.2) and user management (Q3.1.1) actually were available on OJS. This might be because some participants in TG did not have time to access help/documents (Q6.4.2) while some participants in FU were not required to assign roles and had not accessed this feature (Q3.1.1.) in OJS by the time they completed the post-test questionnaire.

IRR:

Within a group, the participants' agreements on the statements of the criteria measuring the features of OJS CL usability in the Collaborativity, Effectiveness and Efficiency dimensions were consistent but their agreements on the statements in Satisfaction dimension were inconsistent.

Both IR & IRR:

Collaborativity and Satisfaction dimensions had many excluded cases/statements in IR and IRR analysis, and their sizes of cases were very small, so the findings about these two dimensions' IR and IRR should be just indicative.

About the importance of the criteria to system usability and its evaluation, the main outcomes from the IR and IRR analysis based on the data collected from Part 2 of Post-test questionnaire are:

IR:

According to both groups' data, the importance levels of the criteria in the Satisfaction dimension were acceptably internally consistent. Removing Q6a4 (*help/documentation*) from the Satisfaction dimension might improve the Satisfaction's internal consistency based on TG's data. The FU's importance levels of the criteria in the Error Tolerance dimension and the TG's importance levels of the criteria in the Collaborativity dimension were internally consistent while both

groups' importance levels of the criteria in the Effectiveness and Efficiency dimensions were internally inconsistent. This indicates that the criteria in the Effectiveness and Efficiency dimensions may need to be either re-tested or restructured.

According to a group's IR analysis, three criteria that were identified to be removed from their dimensions in order to increase their dimensions' internal consistency are "Communication" (Q3a3) and "Security" (Q3a8) in the Collaborativity dimension, and "Help/Documentation" (Q6a4) in the Satisfaction dimension. However the findings in Part 2 of Post-test questionnaire in Table 6-19 (on pg 103) indicate that they are very important or important to CSCL system usability and in fact these three items should be included in the framework.

IRR:

TG's rankings of the importance levels of the criteria in the Satisfaction dimension were excellently consistent, but those in the Collaborativity dimension were fairly consistent and those in the Effectiveness and Efficiency dimensions were not consistent. In contrast, FU's rankings of the importance levels of the criteria in the Effectiveness and Efficiency dimensions were fairly consistent but those in the Collaborativity and Satisfaction dimensions were poor. The reason might be that the participants in FU had more time of using OJS and had gained more experiences in using OJS and working collaboratively in a team than TG's participants had. But they experienced some problems (shown on Appendix 27) when they completed their group assignments so different participants had different thoughts about the importance levels of the criteria in the Satisfaction and Collaborativity dimensions but had similar thoughts on the importance levels of the criteria in the Effectiveness and Effectiveness dimensions. In contrast, most participants in TG were unfamiliar with OJS and had to complete the predefined testing task by following the steps shown on the task script within a short time span. So the participants in TG might have got different ideas about the importance levels of the criteria in the Effectiveness and Efficiency dimensions and had similar opinions about the importance levels of the criteria in the Collaborativity and Satisfaction dimensions.

Table 5-7 Summary of the IR Analysis outputs - Part 1 & Part 2

Post-test Questionnaire		**Part 1 (Agreements on OJS usability)**		**Part 2 (importance levels of Criteria)**	
		Group			
Dimensions	**Measures**	TG	FU	TG	FU
Effectiveness	Case N/Total N	11 / 12	10 /13	12 /12	13/ 13
	IR status	good	acceptable	unacceptable	unacceptable
	Items to be deleted for IR improvement	-	Q1.1.1	-	-
Efficiency	Case N/Total N	7 / 12	9 / 13	12 /12	13 /13
	IR status	good	good	unacceptable	unacceptable
	Items to be deleted for IR improvement	Q2.1.2	Q2.1.2	-	-
Collabora-tivity	Case N/Total N	3 / 12	3 / 13	10 /12	11 /13
	IR status	questionable	unacceptable	good	questionable
	Items dropped by SPSS	Q3.4.2	Q3.1.1, Q3.6.1, Q3.7.1	-	-
	Items to be deleted for IR improvement	Q3.4.1, Q3.6.1, Q3.6.2	Q3.4.1,	Q3a3	Q3a8
Error Tolerance	Case N/Total N	6 / 12	9 / 13	11 /12	10 / 13
	IR status	acceptable	excellent	questionable	good
	Items to be deleted for IR improvement	-	-	-	-
Satisfaction	Case N/Total N	3 / 12	7 /13	10 / 12	13 / 13
	IR status	unacceptable	excellent	acceptable	acceptable
	Items dropped by SPSS	Q6.2.2, Q6.4.2	-	-	-
	Items to be deleted for IR improvement	-	-	Q6a4	-

Note: IR: Internal Reliability;
IR status: mainly based on the values of Cronbach's Alpha;
Items were dropped from the dimension by SPSS because they have zero variance in SPSS. The items are
Q3.1.1 (easy to add /assign users, and manage user roles/accounts on the system),
Q3.4.2 (able to manage my files/notes and the shared files/notes),
Q3.6.1 (able to send a notification to the team after completion of a task),
Q3.7.1 (file sharing and protection),
Q6.2.2 (Tasks can be performed in a straight-forward manner),
Q6.4.2 (easy to access help documents).

Items to be deleted for IR improvement:
Part 1: Q1.1.1 (complete a task on this system within a proper time frame),
Q2.1.2 (The System speed is fast enough);
Q3.4.1 (Being a moderator, I am able to give online instructions, and monitor teamwork on the system),
Q3.6.1 (After I complete a task, I am able to send a notification to the team),
Q3.6.2 (able to find out the status of a task/teamwork)
Part 2: Q3a3 (Communication), Q3a8 (Security), Q6a4 (Help/Documentation)

For the Effectiveness, Efficiency, Collaborativity dimensions, the participants' IRR consistencies were poor or fair. So this would negatively impact the IR reliability of these three dimensions. In other words, the IR result indicating inconsistency in the importance levels of the criteria in these three dimensions might not be reliable. For

example, about Collaborativity dimension, both FU's IR and IRR are inconsistent, this could infer that the actual internal consistency (IR) of the importance of the criteria in the Collaborativity dimension might be better than what its IR analysis's outcomes indicated because of the poor consistency of IRR in FU.

Table 5-8 Summary of the IRR Analysis outputs - Part 1 & Part 2

		Part 1 (Agreements on OJS usability)		Part 2 (Criteria importance)	
IRR Analysis outputs		TG	FU	TG	FU
Effectiveness	Case N/Total N	4 / 6	4 / 6	4 / 4	4 / 4
	IRR status	fair	good	poor	fair
	Items to be deleted for IR improvement	-	-	-	-
Efficiency	Case N/Total N	6 / 8	7 / 8	4 / 4	4 / 4
	IRR status	excellent	fair	poor	fair
	Items to be deleted for IR improvement	-	-	-	-
Collaborativity	Case N/Total N	4 / 13	5 / 13	6 / 8	5 /8
	IRR status	excellent	excellent	fair	poor
	Items to be deleted for IR improvement	-	-	-	-
Satisfaction	Case N/Total N	7 / 12	6 /12	4 / 5	5 / 5
	IRR status	poor	poor	excellent	poor
	Items to be deleted for IR improvement	-	-	-	-

Note: IRR: Inter-Rater Reliability;
Error Tolerance" and "Universal Accessibility" dimensions had too few cases (N = 0), no outputs for them.
IRR status: mainly based on the values of Intraclass Correlation Coefficient

5.5 The concern on the reliability analysis in this study

The main concern of data reliability analysis in this study is the small size of the data sample. A common rule recommended by Field (2005) is that *"a research has at least 10-15 participants per variable"*. So in this case study, the sample size should be 240 at least because there are 24 variables in the Post-test Questionnaire. In fact, the total number of the participants in this case study is 25. So, this study's sample size is far too small. In addition, some participants did not answer all the questions or chose "Not applicable" (N/A), and then their data were excluded from the data reliability analysis in SPSS. This made the number of the cases even smaller. A typical example is the Collaborativity dimension in Part 1's internal reliability analysis. Only three cases' datasets from each group were processed in SPSS. Therefore the outputs of the reliability analysis should be referential, inconclusive, and indicative.

5.6 Summary

This chapter has illustrated the outputs of the Internal Reliability and Inter-Rater Reliability analysis. The Universal Accessibility dimension was not included in the IR analyses while the Universal Accessibility and Error Tolerance dimensions were excluded from the IRR analysis because the numbers of their variables are too small to be processed in SPSS.

The outcomes of IR and IRR analysis illustrate that Part 1 (the agreements on the statements and the criteria referred to the features of OJS CL usability) has better reliability and consistency than Part 2 (the rankings of importance levels of the criteria to system usability and its evaluation). About internal reliability, for Part 1, based on both TG and FU's data, the agreements on the statements of the criteria in the Effectiveness, Efficiency, and Error Tolerance dimensions had internal consistency. For Part 2, both groups' rankings of importance levels of the criteria in the Effectiveness and Efficiency dimensions were inconsistent but their rankings of importance levels of the criteria in the Satisfaction dimension were consistent. Regarding the Inter-rater consistency, for Part 1, each group had good inter-rater consistency in agreeing to the statements on the criteria referred to the features of OJS CL usability in the Effectiveness, Efficiency and Collaborativity dimensions but had poor consistency in agreeing to the statements on the criteria in the Satisfaction dimension. For Part 2, TG had inter-rater inconsistent rankings of importance levels of the criteria in the Effectiveness and Efficiency dimensions but had fairly or excellent inter-rater consistent rankings of the criteria in the Satisfaction and Collaborativity dimensions. FU had opposite results for the four dimensions.

The main concern about the data reliability analysis is the small size of the data sample. The Collaborativity and Satisfaction dimensions had many excluded cases in IR and IRR analysis, and their sizes of cases analysed in SPSS were small, so the outcomes of data reliability analysis in this study should be considered as referential and indicative information rather than as a final conclusion. Further testing by a larger sample of participants needs to be considered in future study.

Chapter 6

6. Findings from the data analysis

The chapter represents the findings of the study in six sections as shown below and then ends with a summary.

6.1 Pre-Test Questionnaire

6.2 Post-test Questionnaire – Part 1, Ranking Statements

6.3 Post-test Questionnaire – Part 2, Ranking criteria

6.4 Post-test Questionnaire – Part 3, Open ended questions

6.5 Task sheet and Observer data collection sheet

6.6 Interview

6.7 Summary

6.1 Pre-Test Questionnaire

There were four demographic questions and 10 questions about a user's previous experience in using CSCL systems. The questionnaire is shown on Appendix 3. The following notes are meant to help with the interpretation of tables presented in this section:

- **%:** percentage of a group who selected a choice about a given question;
- **#:** the number of participants in a group who selected a choice about a given question.

6.1.1 Demographic Questions

Professional Status

In TG, 83.33% (10) participants were students and 16.66% (2) were academic staff whereas in FU, all participants (13) were students. One person in TG was simultaneously a student and a member of staff and a researcher. One person who joined both TG and FU was both a student and a software developer.

Main areas of work/study

All 25 participants across both TG and FU had either worked or studied in computing areas. However, most participants, except 2 persons from TG and FU respectively did not specify their area of work or study. In TG, one participant worked in database and data warehouse areas and the other participant studied networking. In FU, one

participant studied network administration and the other participant studied computer and information science.

Age and Gender of respondents

66.67% of participants (8) in TG and 76.92% of participants (10) in FU were aged 30 years or younger. The majority of participants in both groups were male - 75% (9) in TG and 69.23% (9) in FU.

In summary, all participants in this case study (12 in TG and 13 in FU including 7 persons who joined both groups) had been studying computing or working in computing related jobs. The majority of the participants were students, younger than 30 years old, and male.

6.1.2 Participants' experiences

Previous use of online systems (Q6)

Table 6-1 Responses to Q6

	Groups	
Reason for using Online system	**TG % (#)**	**FU % (#)**
Search information	100 (12)	100(13)
Shopping	83.33 (10)	61.54 (8)
Banking	91.67 (11)	84.62 (11)
Distance learning /Online study	66.67 (8)	53.85 (7)
Online Teaching	16.67 (2)	7.69 (1)
Other, e.g. research, socialising, etc.	16.67 (2) (e.g. facebook, communication, networking)	15.38 (2) (e.g. research, socialising,)

Note: Participants could select more than one answer.

Table 6-1 shows that all participants had used online systems for various purposes. The majority used online systems to search information, shopping, banking, and distance learning or online study. All 25 participants used online systems for searching information. 66.67% (8 persons) in TG and 53.85% (7 persons) in FU had studied online and taken distance learning courses via online system. Four people (two from each group) had used online systems to socialise and communicate with people.

Previous use of CSCL systems in team work/collaborative tasks (Q7 & Q10)

83.33% of participants (10) in TG and all 13 participants in FU had used CSCL systems. 58.33% (7) of TG and 76.92% (10) of FU had used CSCL to complete teamwork.

Types of CSCL systems previously used (Q8)

The majority of the participants (83%) in TG and 100% of FU had previously used Blackboard (see Table 6-2). But none of TG and only 46% (6) of FU had experience in using OJS.

Table 6-2 Responses to Q8

Groups	TG	FU	TG	FU
System name	% participants		# participants	
Blackboard	83.33	100	10	13
Lotus	0	8.00	0	1
Cecil	0	0.00	0	0
OJS	0	46.00	0	6
Other	25% e.g. Moodle and PingPong; 8.33% no response;	15.00% (Moodle, Webspace)	3	2

Note*:* Respondents could select more than one answer.

Previous use of OJS (Q11 & 13)

16.67% (2) participants in TG and 61.54% (8) participants in FU had previously worked on OJS system, and 8.33% (1) of TG and 53.85% (7) of FU had used OJS for team work. So, FU had a higher percentage of participants, who had worked on a CSCL system and OJS, and had used OJS to complete collaborative tasks/ teamwork, than TG had. These figures are different from those shown on Table 6-2. The figures from Q11 & Q13 should be more accurate as an academic staff in TG had used OJS before and the students in FU had started working on their group assessments on OJS at that time when they completed the questionnaires.

Time spent on using online/CSCL/OJS systems (Q5/Q9/Q12)

Table 6-3 below shows that all participants from FU and most from TG had the experience in using online systems. 91.67% (11 participants) of TG and 84.52% (11) of FU had used online systems for 4 years or longer. The majority of FU (69.23%) and 41.67% of TG had used a CSCL system for less than a year. None of FU but 24.99% (3) in TG had used CSCL for more than 2 years.

16.67% (2) of TG participants and 53.85% (7) of FU participants spent less than an hour per week on OJS. 53.85% (7) of TG did not spend any time on OJS. Both groups had similar percentages of users (15-16% or 2 people) who had spent 1-4 hours on OJS per week.

Table 6-3 Responses to Q5, Q9, &Q12

Question 5: Using online systems						
	Time					
	NA	**≤ 1yrs**	**2yrs**	**3yrs**	**4yrs**	**>4yrs**
Groups	% (#)	% (#)	% (#)	% (#)	% (#)	% (#)
TG	8.33 (1)	0 (0)	0 (0)	0 (0)	25 (3)	66.67 (8)
FU	0 (0)	7.69 (1)	0 (0)	7.69 (1)	7.69 (1)	76.92 (10)
Question 9: Use of CSCL systems						
	Time					
	NA /NR	**≤ 1yrs**	**2yrs**	**3yrs**	**4yrs**	**>4yrs**
Groups	% (#)	% (#)	% (#)	% (#)	% (#)	% (#)
TG	16.67(2)/ 8.33(1)	41.67 (5)	8.33 (1)	8.33 (1)	8.33 (1)	8.33 (1)
FU	0 (0)	69.23 (9)	30.77 (4)	0 (0)	0 (0)	0 (0)
Question 12: weekly use of OJS						
	Time					
	NA (NR)	Nil	≤ 1hr	1 ≤4hrs	4 ≤ 10hrs	>10hrs
Groups	% (#)	% (#)	% (#)	% (#)	% (#)	% (#)
TG	8.33 (1)	58.33 (7)	16.67 (2)	16.67 (2)	0 (0)	0 (0)
FU	7.69 (1)	23.08 (3)	53.85 (7)	15.38 (2)	0 (0)	0 (0)

Note: **NA:** Not Applicable. **NR**: Not Response;

Rating CSCL expertise

Table 6-4 Responses to Q14

	Expertise Level					
	1 Novice	2 Beginner	3 Intermediate	4 Advanced	5 Expert	Mean
Groups	% (#)	% (#)	% (#)	% (#)	% (#)	
TG	33.33 (4)	41.67 (5)	8.33 (1)	16.67 (2)	0 (0)	2.08
FU	30.77 (4)	23.08 (3)	38.46 (5)	7.69 (1)	0 (0)	2.23

Table 6-4 shows that in TG, 41.67% of participants (5) rated themselves as level 2 (beginner) whereas in FU, 38.46% (5) rated themselves at level 3 (intermediate). At level 4 (advanced), TG has 2 participants while FU had 1 participant. The TG's average level is 2.08, a bit lower than FU's, which is 2.23.

6.1.3 Summary

The majority of participants believed that their CSCL expertise levels were either level 1 (Novice) or 2 (Beginner). Most participants had used online systems for four years or longer but few participants had worked on CSCL systems (e.g. Blackboard) for four years or longer. Before doing testing and completing the group assignment, in TG only 2 participants had ever used OJS while only one person had used OJS for team work. Over 75% of TG and FU spent less than an hour or did not spend any time per week on using OJS.

6.2 Post-Test Questionnaire – Part 1: Ranking Statements

Part 1 of the Post-test questionnaire comprised 43 statements that were ranked according to a 5 point Likert scale (see Appendix 4). A statement states a key feature of OJS CL usability measured by one of 25 criteria which are grouped into six dimensions. The data was collected from two groups (TG and FU). Each group's data was analysed in SPSS and the outputs for both TG and FU are listed in Appendix 23 and Appendix 24. The findings are presented in sections 6.2.1 – 6.2.6. The following notes are intended to help with interpretation of the tables and figures presented in following sections.

Notes for Tables:

- % - percentage of participants
- # - number of participants

Notes for Figures

- The upper edge of the box indicates the average level of agreement for a given statement
- The lower edge of the box indicates the bottom 25th percentile
- The top and bottom whiskers show the maximum level and the minimum level of an agreement respectively for a given statement within a group.

6.2.1 Effectiveness Dimension

Effectiveness is the first dimension to be analysed. It consists of four criteria and six statements. The rankings from both groups are shown in Table 6-5 below. 67% of TG (8 participants) were *undecided* about statement 1.3.1 while 54% of FU (7 participants) agreed with Q1.3.2. Nobody in TG and FU selected *N/A.* One participant from each group did not respond to Q1.3.2, and one participant in TG and three participants in FU did not respond to Q1.4.2.

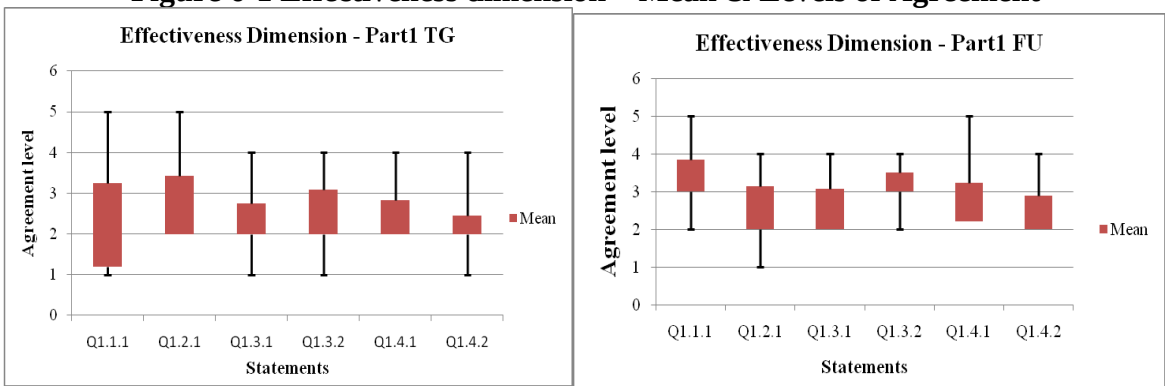
FU has a higher overall average than TG's (see Figure 6-1 below). TG's Mean range is between 2.45 (for Q1.4.2) and 3.42 (for Q1.2.1) i.e. from *disagree* to *undecided* while FU's is between 2.90 (for Q1.4.2) and 3.85 (for Q1.1.1), i.e. from *near to undecided* to *near to agreed.*

Table 6-5 Effectiveness Dimension – ranking statements

Frequency Percentages/Counts Comparison between TG and FU

Level of Agreement		Not Applicable (N/A)		No Response (N/R)		Strongly Disagree 1		Disagree 2		Undecided 3		Agree 4		Strongly Agree 5	
%/#		%/#		%/#		%/#		%/#		%/#		%/#		%/#	
	Groups	TG	FU	TG	FU	TG	FU	TG	FU	TG	FU	TG	FU	TG	FU
Statements on OJS	Q1.1.1	0/0	0/0	0/0	0/0	17/2	0/0	0/0	8/1	33/4	23/3	42/5	46/6	8/1	23/3
	Q1.2.1	0/0	0/0	0/0	0/0	0/0	8/1	25/3	15/2	25/3	31/4	33/4	46/6	17/2	0/0
	Q1.3.1	0/0	0/0	0/0	0/0	8/1	0/0	17/2	23/3	67/8	46/6	8/1	31/4	0/0	0/0
	Q1.3.2	0/0	0/0	8/1	8/1	8/1	0/0	17/2	8/1	25/3	31/4	42/5	54/7	0/0	0/0
	Q1.4.1	0/0	0/0	0/0	0/0	0/0	8/1	42/5	8/1	33/4	46/6	25/3	31/4	0/0	8/1
	Q1.4.2	0/0	0/0	8/1	23/3	8/1	0/0	42/5	31/4	33/4	23/3	8/1	23/3	0/0	0/0

Figure 6-1 Effectiveness dimension – Mean & Levels of Agreement



6.2.2 Efficiency Dimension

Efficiency dimension has four criteria and eight statements. The rankings from both groups are presented in Table 6-6 below. Q2.1.1 and Q2.1.2 were each ranked at *agree* by 6 participants (50%), Q2.2.1 and Q2.4.3 were ranked at *undecided* by 7 participants (58%) and 6 participants (50%) respectively in TG; in FU, Q2.1.2 was ranked at *agree* by 62% (8 participants). All participants responded to the statements. 42% of TG (5 participants) and 31% of FU (4 participants) selected *N/A* for Q2.3.1. Only one person in TG selected *N/A* for Q2.4.1.

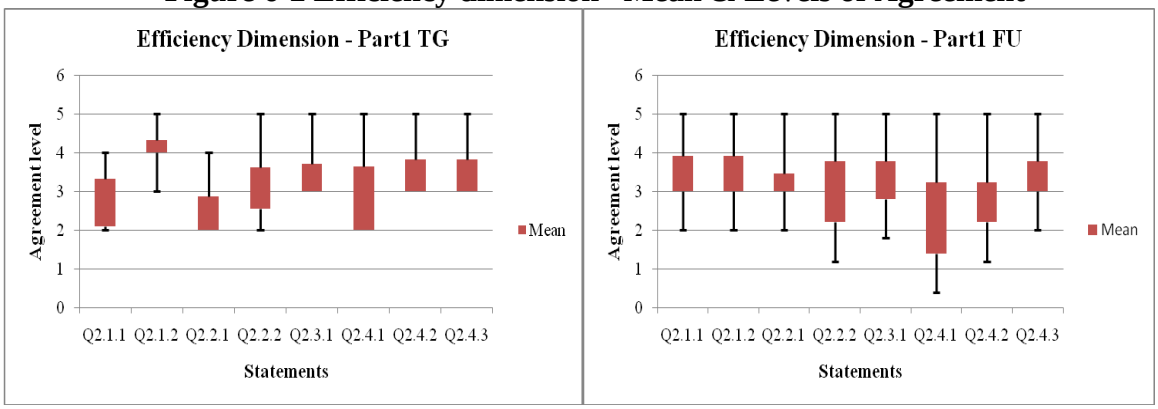
The average agreement levels are higher in TG than in FU. TG's range is between 2.88 (for Q2.2.1) and 4.33 (for Q2.1.2), from *undecided* to *agree* while FU's is between 3.23 (for Q2.4.1 & Q2.4.2) and 3.92 (for Q2.1.1 & Q2.1.2), from *undecided to near to agree* (see Figure 6-2 below).

Table 6-6 Efficiency dimension – ranking statements

Frequency Percentages/Counts Comparison between TG and FU

Level of Agreement		Not Applicable (N/A)		No Response (N/R)		Strongly Disagree 1		Disagree 2		Undecided 3		Agree 4		Strongly Agree 5	
%/#		%/#		%/#		%/#		%/#		%/#		%/#		%/#	
Groups		TG	FU	TG	FU	TG	FU	TG	FU	TG	FU	TG	FU	TG	FU
Statements on OJS	Q2.1.1	0/0	0/0	0/0	0/0	0/0	0/0	17/2	0/0	33/4	31/4	50/6	46/6	0/0	23/3
	Q2.1.2	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	8/1	23/3	50/6	62/8	42/5	15/2
	Q2.2.1	0/0	0/0	0/0	0/0	0/0	0/0	25/3	8/1	58/7	46/6	17/2	38/5	0/0	8/1
	Q2.2.2	0/0	0/0	0/0	0/0	0/0	0/0	8/1	15/2	41/5	23/3	25/3	31/4	25/3	31/4
	Q2.3.1	42/5	31/4	0/0	0/0	0/0	0/0	0/0	8/1	25/3	15/2	25/3	31/4	8/1	15/2
	Q2.4.1	8/1	0/0	0/0	0/0	0/0	15/2	17/2	0/0	25/3	38/5	25/3	38/5	25/3	8/1
	Q2.4.2	0/0	0/0	0/0	0/0	0/0	8/1	0/0	8/1	42/5	46/6	33/4	31/4	25/3	8/1
	Q2.4.3	0/0	0/0	0/0	0/0	0/0	0/0	0/0	8/1	50/6	31/4	17/2	38/5	33/4	23/3

Figure 6-2 Efficiency dimension– Mean & Levels of Agreement



6.2.3 Collaborativity dimension

Collaborativity dimension consists of eight criteria and 13 statements. The rankings from both groups are presented in Table 6-7 below. Q3.5.1, Q3.8.2, Q3.6.1, and Q3.6.2 were ranked at *agree* by 6 participants (50%), 6 participants (50%), 7 participants (58%), and 8 participants (67%) respectively in TG; Q3.3.1 was graded at *undecided* by 6 participants (50%) in TG; Q3.6.2 was ranked at *agree* by 8 participants (62%) in FU. Several statements were ranked at *N/A*, e.g. in FU, Q3.1.1 was ranked at N/A by 8 participants (62%), Q3.1.2 by 7 participants (54%); in TG, Q3.7.1 by 8 participants (67%), and Q3.8.1 by 6 participants (50%).

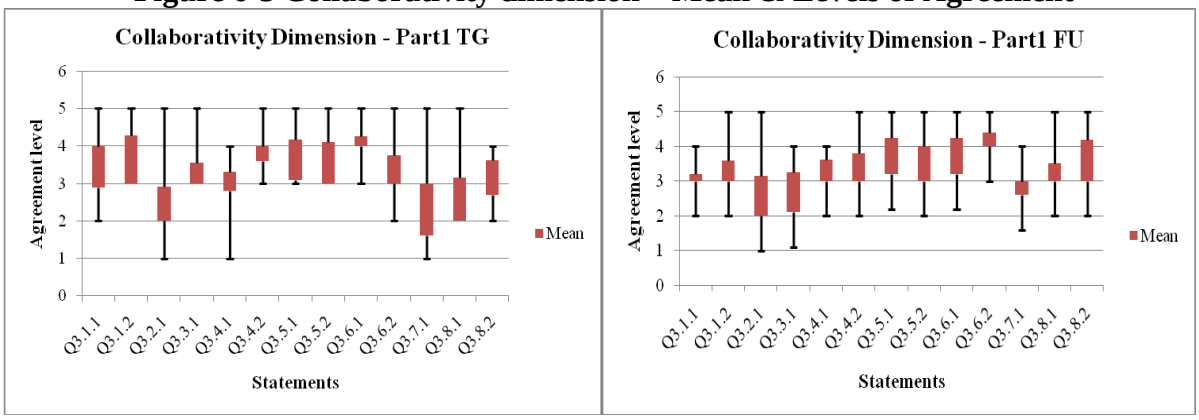
Figure 6-3 below shows that FU's average agreement levels are similar to TG's. TG's range is between 2.92 (for Q3.2.1) and 4.27 (for Q3.1.2) while FU's is between 3.00 (for Q3.7.1) and 4.38 (for Q3.6.2). All responses were dispersed from *undecided* to *agree* except the responses to Q3.2.1 from TG whose Mean is lower than 3 (*undecided* level).

Table 6-7 Collaborativity dimension– ranking statements

Frequency Percentages/Counts Comparison between TG and FU

Level of Agreement		Not Applicable (N/A)		No Response (N/R)		Strongly Disagree 1		Disagree 2		Undecided 3		Agree 4		Strongly Agree 5	
%/#		%/#		%/#		%/#		%/#		%/#		%/#		%/#	
Groups		TG	FU	TG	FU	TG	FU	TG	FU	TG	FU	TG	FU	TG	FU
Statements on OJS	Q3.1.1	17/2	62/8	0/0	0/0	0/0	0/0	8/1	0/0	8/1	31/4	42/5	8/1	25/3	0/0
	Q3.1.2	8/1	54/7	0/0	8/1	0/0	0/0	0/0	0/0	17/2	23/3	33/4	8/1	42/5	8/1
	Q3.2.1	0/0	0/0	0/0	0/0	8/1	8/1	33/4	23/3	25/3	23/3	25/3	38/5	8/1	8/1
	Q3.3.1	8/1	8/1	0/0	0/0	0/0	0/0	0/0	15/2	50/6	38/5	33/4	38/5	8/1	0/0
	Q3.4.1	17/2	38/5	0/0	0/0	8/1	0/0	0/0	0/0	33/4	23/3	42/5	38/5	0/0	0/0
	Q3.4.2	33/4	8/1	8/1	15/2	0/0	0/0	0/0	0/0	8/1	23/3	42/5	46/6	8/1	8/1
	Q3.5.1	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	17/2	15/2	50/6	46/6	33/4	38/5
	Q3.5.2	25/3	0/0	0/0	0/0	0/0	0/0	0/0	8/1	17/2	23/3	33/4	31/4	25/3	38/5
	Q3.6.1	0/0	0/0	0/0	0/0	0/0	0/0	0/0	8/1	8/1	8/1	58/7	38/5	33/4	46/6
	Q3.6.2	0/0	0/0	0/0	0/0	0/0	0/0	8/1	0/0	17/2	0/0	67/8	62/8	8/1	38/5
	Q3.7.1	67/8	46/6	0/0	0/0	8/1	0/0	0/0	8/1	17/2	38/5	0/0	8/1	8/1	0/0
	Q3.8.1	50/6	8/1	0/0	0/0	0/0	8/1	17/2	0/0	17/2	46/6	8/1	15/2	8/1	23/3
	Q3.8.2	33/4	15/2	0/0	8/1	0/0	0/0	8/1	0/0	8/1	15/2	50/6	31/4	0/0	31/4

Figure 6-3 Collaborativity dimension – Mean & Levels of Agreement



6.2.4 Error Tolerance Dimension

The criterion error rate was not included and recorded during testing sessions as it was not the goal of the testing. When errors occurred during the testing sessions, most participants took a few seconds/minutes to find the ways to resolve them because the errors were often related to unfamiliarity, i.e. the participants were unfamiliar with OJS, for example, clicking a wrong link or wrong button. So, in this dimension, the data about one criterion (Error Prevention) and its three statements were processed.

Table 6-8 below shows the rankings from both groups. None of the statements was ranked at an agreement level by over 50% of participants in a group. All participants responded to the statements. The three statements were ranked at *N/A* by some participants, e.g. in TG, Q4.2.2 by 5 participants (42%), Q4.2.3 by 4 participants (33%) and Q4.2.1 by 3 participants; in FU, Q4.2.3 by 3 participants (23%), Q2.4.1 and Q4.2.2

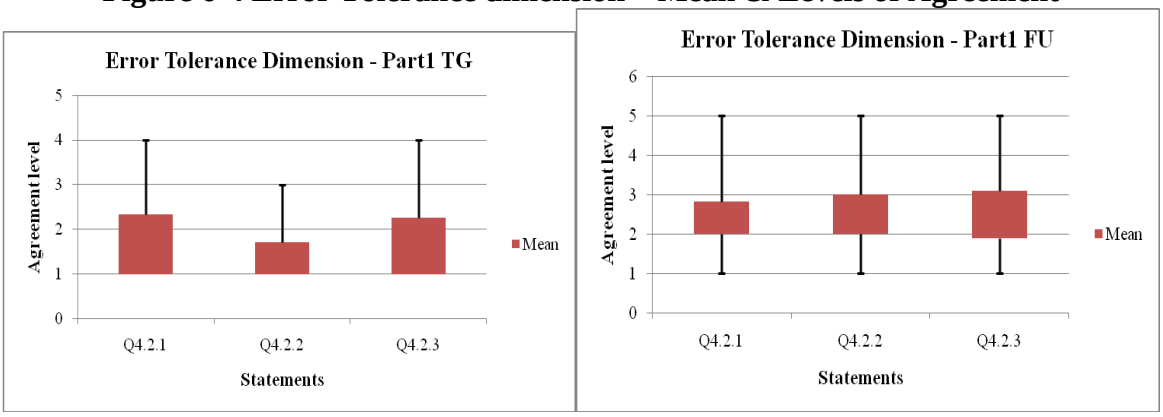
by 2 participants (15%) respectively. The two groups' average agreement levels are quite different. FU's range is higher than TG's. TG's is between 1.71 (for Q4.2.2) and 2.33 (for Q4.2.1), i.e. between *strongly disagree* but *near to disagree* and *disagree* but FU's is between 2.82 (for Q4.2.1) and 3.1 (for Q4.2.3), i.e. *around undecided* (see Figure 6-4).

Table 6-8 Error Tolerance Dimension– ranking statements

Frequency Percentages/Counts Comparison between TG and FU

Level of Agreement		Not Applicable (N/A)		No Response (N/R)		Strongly Disagree 1		Disagree 2		Undecided 3		Agree 4		Strongly Agree 5	
%/#		%/#		%/#		%/#		%/#		%/#		%/#		%/#	
Groups		TG	FU	TG	FU	TG	FU	TG	FU	TG	FU	TG	FU	TG	FU
Statement on OJS	Q4.2.1	25/3	15/2	0/0	0/0	33/4	8/1	8/1	23/3	8/1	38/5	25/3	8/1	0/0	8/1
	Q4.2.2	42/5	15/2	0/0	0/0	33/4	8/1	8/1	23/3	17/2	23/3	0/0	23/3	0/0	8/1
	Q4.2.3	33/4	23/3	0/0	0/0	25/3	8/1	8/1	15/2	25/3	31/4	8/1	8/1	0/0	15/2

Figure 6-4 Error Tolerance dimension – Mean & Levels of Agreement



6.2.5 Universal Accessibility (Ubiquity) Dimension

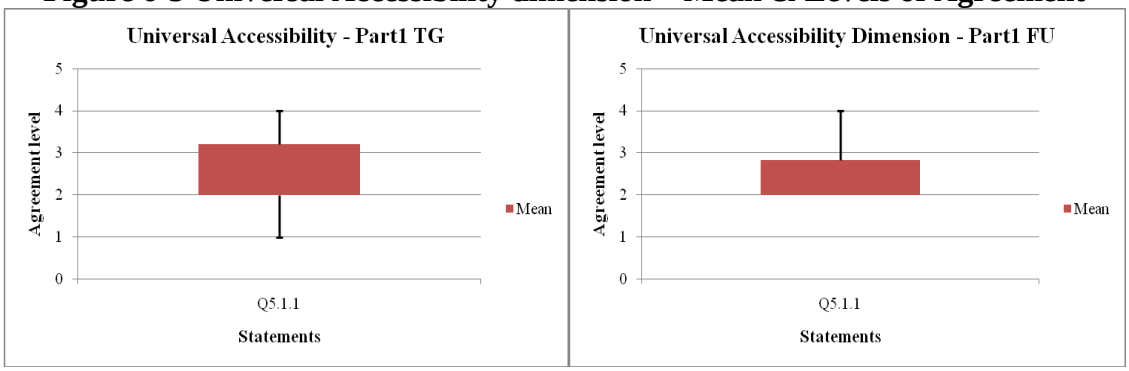
Universal Accessibility dimension has one criterion and one statement. Table 6-9 below depicts that Q5.1.1 was ranked at *agree* by 42% of TG (5 participants), at *disagree* by 38% of FU (5 participants). Each group had one participant who ranked the statement as *N/A*. TG's average agreement level is 3.2, i.e. *undecided*, higher than FU's, which is 2.82, *near to undecided* (see Figure 6-5 below).

Table 6-9 Universal Accessibility Dimension – ranking statements

Frequency Percentages/Counts Comparison between TG and FU

Level of Agreement	Not Applicable (N/A)		No Response (N/R)		Strongly Disagree 1		Disagree 2		Undecided 3		Agree 4		Strongly Agree 5	
%/#	%/#		%/#		%/#		%/#		%/#		%/#		%/#	
Groups	TG	FU	TG	FU	TG	FU	TG	FU	TG	FU	TG	FU	TG	FU
Q5.1.1	8/1	8/1	8/1	8/1	0/0	0/0	25/3	38/5	17/2	23/3	42/5	23/3	0/0	0/0

Figure 6-5 Universal Accessibility dimension – Mean & Levels of Agreement



6.2.6 Satisfaction Dimension

Satisfaction dimension consists of six criteria and 12 statements. The rankings from both groups are illustrated on Table 6-10. In TG, Q6.1.1 and Q6.2.2 were both ranked at *agree* by 8 participants (67%), Q6.1.2 and Q6.6.2 at *agree* by 7 participants (58%) respectively, Q6.6.1 at *undecided* by 6 participants (50%); in FU, Q6.1.2 was ranked at *agree* by 7 persons (54%). More participants in TG than in FU selected *N/A* for some statements, e.g. in TG, Q6.4.2 and Q6.4.3 were each ranked at *N/A* by 9 participants (75%), Q6.6.1 by 4 participants (33%) while in FU, Q6.4.3 was ranked at *N/A* by 3 participants (23%) and Q6.6.1 by 1 participant (8%).

Table 6-10 Satisfaction Dimension– ranking statements

Frequency Percentages/Counts Comparison between TG and FU

Level of Agreement		Not Applicable (N/A)		No Response (N/R)		Strongly Disagree 1		Disagree 2		Undecided 3		Agree 4		Strongly Agree 5	
%/#		%/#		%/#		%/#		%/#		%/#		%/#		%/#	
	Groups	TG	FU	TG	FU	TG	FU	TG	FU	TG	FU	TG	FU	TG	FU
Statements on OJS	Q6.1.1	0/0	0/0	0/0	0/0	0/0	8/1	0/0	0/0	25/3	38/5	67/8	46/6	8/1	8/1
	Q6.1.2	0/0	0/0	0/0	0/0	0/0	0/0	8/1	0/0	25/3	38/5	58/7	54/7	8/1	8/1
	Q6.1.3	17/2	0/0	0/0	0/0	0/0	0/0	0/0	15/2	33/4	46/6	42/5	31/4	8/1	8/1
	Q6.2.1	0/0	0/0	0/0	0/0	0/0	8/1	17/2	23/3	25/3	8/1	42/5	31/4	17/2	31/4
	Q6.2.2	0/0	0/0	0/0	8/1	0/0	0/0	8/1	23/3	17/2	23/3	67/8	31/4	8/1	15/2
	Q6.3.1	0/0	0/0	0/0	8/1	0/0	8/1	25/3	8/1	25/3	15/2	33/4	31/4	17/2	31/4
	Q6.4.1	42/5	0/0	0/0	0/0	0/0	0/0	8/1	23/3	33/4	46/6	17/2	23/3	0/0	8/1
	Q6.4.2	75/9	15/2	0/0	0/0	0/0	0/0	0/0	0/0	25/3	46/6	0/0	31/4	0/0	8/1
	Q6.4.3	75/9	23/3	0/0	0/0	0/0	0/0	0/0	0/0	17/2	46/6	8/1	23/3	0/0	8/1
	Q6.5.1	0/0	0/0	0/0	8/1	0/0	8/1	33/4	15/2	33/4	31/4	17/2	23/3	17/2	15/2
	Q6.6.1	33/4	8/1	0/0	0/0	0/0	8/1	0/0	8/1	50/6	31/4	17/2	38/5	0/0	8/1
	Q6.6.2	0/0	0/0	0/0	0/0	0/0	0/0	17/2	15/2	25/3	31/4	58/7	46/6	0/0	8/1

Figure 6-6 Satisfaction dimension – Mean & Levels of Agreement

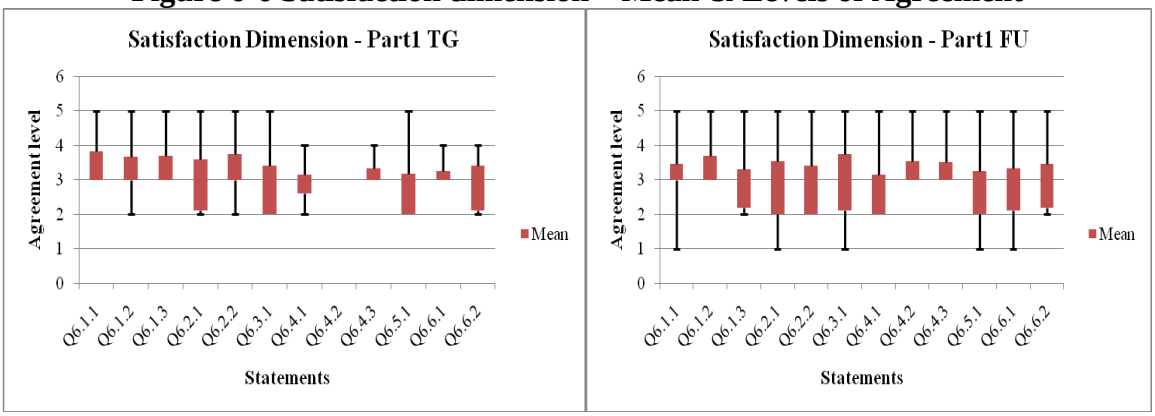


Figure 6-6 shows that the two groups' average agreement levels are quite similar, all are over 3 but less than 4. TG's range is between 3 (for Q6.4.2) (*undecided*) and 3.83 (for Q6.1.1) (*undecided* but *near to agree*) while FU's is between 3.15 (for Q6.4.1) (*undecided*) and 3.75 (for Q6.3.1) (*undecided* but *near to agree*).

6.2.7 Summary

Table 6-11 next page summarises the findings on the six tables 6-5 - 6-10 and shows that what statements were ranked at a particular level by 50% or more participants in a group. 23 statements were agreed or strongly agreed, 5 statements were undecided, and one statement was strongly disagree and disagreed by 50% or more of TG while 17 statements were agreed or strongly agreed by 50% or more of FU. Overall, 16 out of 43 statements (37.21%) highlighted in Turquoise colour in Table 6-11 were strongly agreed or agreed by more than 50% of each group. Among these 16 statements, 6 statements are included in the Collaborativity dimension, 5 statements in the Satisfaction dimension, 4 in the Efficiency dimension, and 1 statement in the Effectiveness dimension.

Table 6-12 on pg 97 listed 20 statements which were ranked at N/A by some participants from TG or FU. Six statements that were ranked at N/A by 50% or more of a group and are highlighted in Turquoise colour are Q6.4.2 and Q6.4.3 (the statements for the criterion *Help/Documentation)*, Q3.7.1 (the statement for *File/Content Protection*), Q3.8.1 (the statement for *Security*), and Q3.1.1 and Q3.1.2 (the statements for *User Management*).

Both TG and FU's average ranges for the five dimensions - Effectiveness, Efficiency, Collaborativity, Universal Accessibility, and Satisfaction are between *near to undecided*

to *agree*. For Error Tolerance, the average ranges are lower than the ranges for the other five dimensions. Its Mean ranges are between *near to disagreed* to *disagreed* in TG and between *near to undecided* to *undecided* in FU.

Table 6-11 The statements ranked by 50% or more participants in a group

			Rankings			
		Criteria (Statements)	**Not applicable**	**Strongly disagree & Disagree**	**Undecided**	**Strongly agree &Agree**
Dimensions	Effectiveness	Completeness (1.1.1)				TG, FU
		Visibility (1.2.1)				TG
		Organisation/Design (1.3.1 - User Friendly/Familiar)			TG	
		Organisation/Design (1.3.2 - Logical Steps)				FU
		Navigability (1.4.2 - clearly points me to the next step/task in a workflow)		TG		
	Efficiency	Speed (2.1.1 - Work on tasks efficiently)				TG, FU
		Speed (2.1.2 - Speed of system)				TG, FU
		Familiarity/Consistency/Standards (2.2.1 - Familiar user interface)			TG	
		Familiarity/Consistency/Standards (2.2.2 - Consistent Layout)				TG, FU
		Effort (2.4.1 - no need to remember information)				TG
		Effort (2.4.2 - no need for prior knowledge)				TG
		Effort (2.4.3 - no need for technical support)			TG	TG, FU
	Collaborativity	User Management (3.1.1 - Manage user account)	FU			TG
		User Management (3.1.2 - Editor able to assign jobs)	FU			TG
		Communication (3.3.1)			TG	
		User Control/Moderator & Teacher Control (3.4.2 - User control)				TG, FU
		File content sharing & management (3.5.1 - easy file upload)				TG, FU
		File content sharing & management (3.5.2 - easy file retrieval)				TG, FU
		Process Tracking/Automated Notification (3.6.1 - notification of task)				TG, FU
		Process Tracking/Automated Notification (3.6.2 - show task status)				TG, FU
		File/Content Protection (3.7.1 - warning message for file sharing)	TG			
		Security (3.8.1 - secure for storing teams' work/files)	TG			
		Security (3.8.2 - logon required)				TG, FU
	Satisfaction	Usefulness/Functionality (6.1.1 - system is fully functional)				TG, FU
		Usefulness/Functionality (6.1.2 - useful for teamwork)				TG, FU
		Usefulness/Functionality (6.1.3 - integrated functions)				TG
		Learnability/Predictability/Recognition/Memorability (6.2.1 - easy to learn)				TG, FU
		Learnability/Predictability/Recognition/Memorability (6.2.2 - task easy to perform)				TG
		Simplicity (6.3.1 - system is simple to use)				TG, FU
		Help/Documentation (6.4.2 - easy to access help documents)	TG			
		Help/Documentation (6.4.3 - easy to switch between help function and work)	TG			
		Overall (6.6.1 - Reliable system)			TG	
		Overall (6.6.2 - Satisfaction with the system)				TG, FU
		Totals number of statements	2 (FU) 4 (TG)	1 (TG)	5 (TG)	17 (FU) 23 (TG)

Note: a statement strongly agreed or agreed by >50% of each group is highlighted in Turquoise colour.

Table 6-12 The statements ranked at "Not Applicable" (N/A)

Q#	**Criteria** **Statements on the features of OJS usability**	% (#) of TG	% (#) of FU
Efficiency dimension			
Q2.3.1	*Flexibility/Adaptability/Configurability* As a site administrator/editor, I can easily modify/configure forms or templates e.g. email templates provided by the system as necessary.	42 (5)	31 (4)
Q2.4.1	*Effort* I don't have to continue remembering information throughout several actions.	8 (1)	0 (0)
Collaborativity dimension			
Q3.1.1	*User Management* It is easy to add/assign users, and manage user roles/accounts on the system.	17 (2)	62 (8)
Q3.1.2	*User Management* Being an Editor, I am able to assign jobs to the teammates.	8 (1)	54 (7)
Q3.3.1	*Communication* I am able to communicate with the teammates or other users on the system as necessary.	8 (1)	8 (1)
Q3.4.1	*User Control/Moderator & Teacher control* Being a moderator, I am able to give online instructions, and monitor teamwork on the system.	17 (2)	38 (5)
Q3.4.2	*User Control/Moderator & Teacher control* As a user, I am able to manage my files/notes and the shared files/notes.	33 (4)	8 (1)
Q3.5.2	*File/Content Sharing & Management* Files can be retrieved easily in the share workspace on the system.	25 (3)	
Q3.7.1	*File/Content Protection* The system would give me a warning when I try modifying files or notes on the share workspace while my teammates are working on them.	67 (8)	46 (6)
Q3.8.1	*Security* The system seems secure for storing teams' work/files.	50 (6)	8 (1)
Q3.8.2	*Security* Users need to logon to modify their artifacts or contact their teammates on the system.	33 (4)	15 (2)
Error tolerance dimension			
Q4.2.1	*Error Prevention* The system warns me if I am about to make a potential error.	25 (3)	15 (2)
Q4.2.2	*Error Prevention* The system gives me error alerts that clearly tell me how to correct errors.	42 (5)	15 (2)
Q4.2.3	*Error Prevention* Whenever I make a mistake, I am able to recover it easily and quickly e.g. by using an "undo" or "cancel" or "reverse" button.	33 (4)	23 (3)
Universal Accessibility			
Q5.1.1	*Support different users with different levels of IT expertise* The system supports both novice and expert users, advance features are available to expert users.	8 (1)	8 (1)
Satisfaction dimension			
Q6.1.3	*Usefulness/Functionality* The various functions in this system are well integrated.	17 (2)	0 (0)
Q6.4.1	*Help/Documentation* The information (such as online help, on-screen messages, and other documentation) provided on this system is clear, understandable, and helpful.	42 (5)	0 (0)
Q6.4.2	*Help/Documentation* It is easy to access help documents.	75 (9)	15 (2)
Q6.4.3	*Help/Documentation* I can easily switch between help and my work.	75 (9)	23 (3)
Q6.6.1	*Overall* The system always is reliable.	33 (4)	8 (1)

Note: a statement ranked as *N/A* by >50% of a group is highlighted in Turquoise colour.

6.3 Post-test Questionnaire – Part 2, Ranking criteria

The participants were asked to use four-point Likert scale to rank an importance level for each criterion to indicate how important they considered the criterion to a CSCL system usability and UE according to their previous experiences in using CSCL systems to complete collaborative tasks but not just according to the experience in using or testing OJS. Similarly to section 6.2, the data collected from two groups was analysed in SPSS and the outputs for each group are presented in Appendix 25 and Appendix 26. The findings are presented in the following sections 6.3.1 to 6.3.6.

In the tables, # presents number of participants, and % means a percentage of participants who ranked an important level of a criterion to a CSCL system usability.

In the figures, the upper edge of a box indicates the average of the importance levels of a criterion, the lower edge of the box indicates the 25th percentile, the top and bottom whiskers show the maximum and minimum importance levels of a criterion ranked by the participants in a group.

6.3.1 Effectiveness Dimension

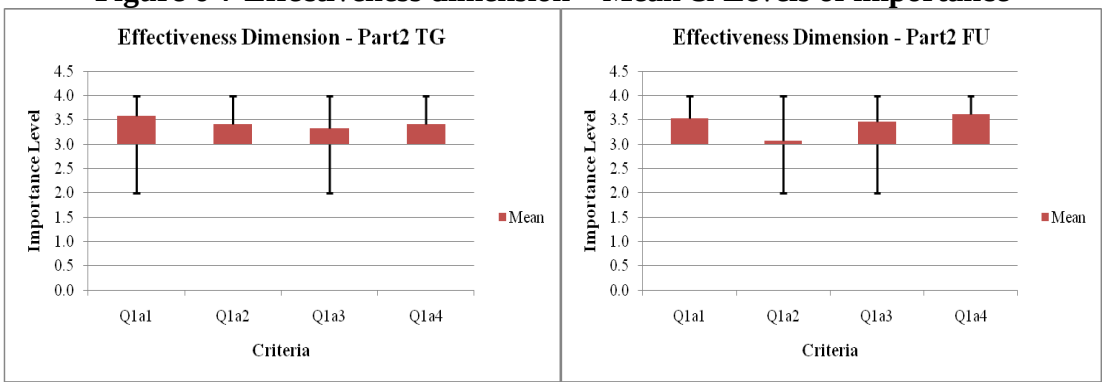
The Effectiveness dimension has four criteria. The rankings from both groups are shown on Table 6-13. All criteria were ranked at either *very important* or *important* by over 50% of participants in a group. Nobody selected *N/A* and *Not important*. The two groups' average importance levels are very similar, at *important level*. TG's range is between 3.33 (for Q1a3) and 3.58 (for Q1a1) while FU's is between 3.08 (for Q1a2) and 3.62 (for Q1a4) (see Figure 6-7).

Table 6-13 Effectiveness Dimension– ranking criteria

Frequency Percentages/Counts Comparison between TG and FU

Level of importance		**Not Applicable (N/A)**		**No Response (N/R)**		**Not important 1**		**A little important 2**		**Important 3**		**Very important 4**	
%/#		%/#		%/#		%/#		%/#		%/#		%/#	
	Groups	TG	FU	TG	FU	TG	FU	TG	FU	TG	FU	TG	FU
Criteria	Q1a1	0/0	0/0	0/0	0/0	0/0	0/0	8/1	0/0	25/3	46/6	67/8	54/7
	Q1a2	0/0	0/0	0/0	0/0	0/0	0/0	0/0	8/1	58/7	77/10	42/5	15/2
	Q1a3	0/0	0/0	0/0	0/0	0/0	0/0	8/1	8/1	50/6	38/5	42/5	54/7
	Q1a4	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	58/7	38/5	42/5	62/8

Figure 6-7 Effectiveness dimension – Mean & Levels of importance



6.3.2 Efficiency Dimension

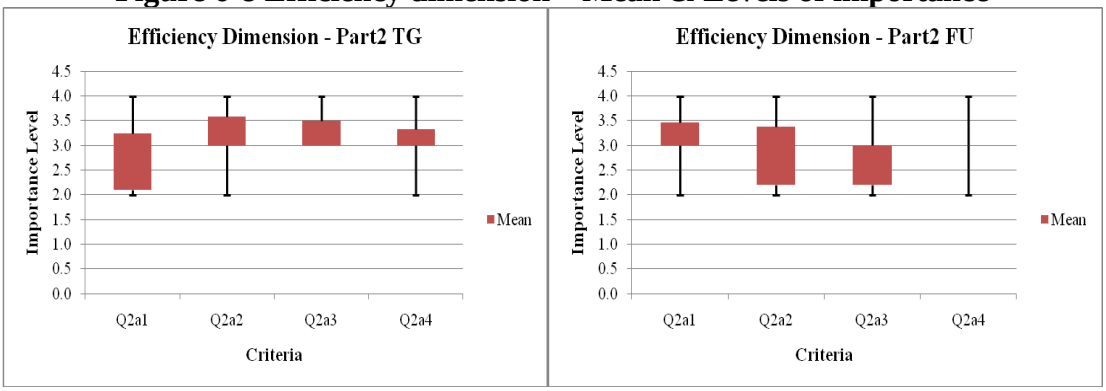
The Efficiency dimension has four criteria. Table 6-14 shows that all criteria were ranked at either *very important* or *important* by over 50% of participants in a group. The two groups' average importance levels are very similar, at *important level*. TG's range is between 3.25 (for Q2a1) and 3.58 (for Q2a2) while FU's is between 3.00 (for Q2a3 &Q2a4) and 3.46 (for Q2a1) (see Figure 6-8).

Table 6-14 Efficiency Dimension– ranking criteria

Frequency Percentages/Counts Comparison between TG and FU

Level of importance		**Not Applicable (N/A)**		**No Response (N/R)**		**Not important 1**		**A little important 2**		**Important 3**		**Very important 4**	
%/#		%/#		%/#		%/#		%/#		%/#		%/#	
	Groups	TG	FU	TG	FU	TG	FU	TG	FU	TG	FU	TG	FU
Criteria	Q2a1	0/0	0/0	0/0	0/0	0/0	0/0	17/2	8/1	42/5	38/5	42/5	54/7
	Q2a2	0/0	0/0	0/0	0/0	0/0	0/0	8/1	15/2	25/3	31/4	67/8	54/7
	Q2a3	0/0	0/0	0/0	0/0	0/0	0/0	0/0	15/2	50/6	69/9	50/6	15/2
	Q2a4	0/0	0/0	0/0	0/0	0/0	0/0	8/1	8/1	50/6	85/11	42/5	8/1

Figure 6-8 Efficiency dimension – Mean & Levels of importance



6.3.3 Collaborativity Dimension

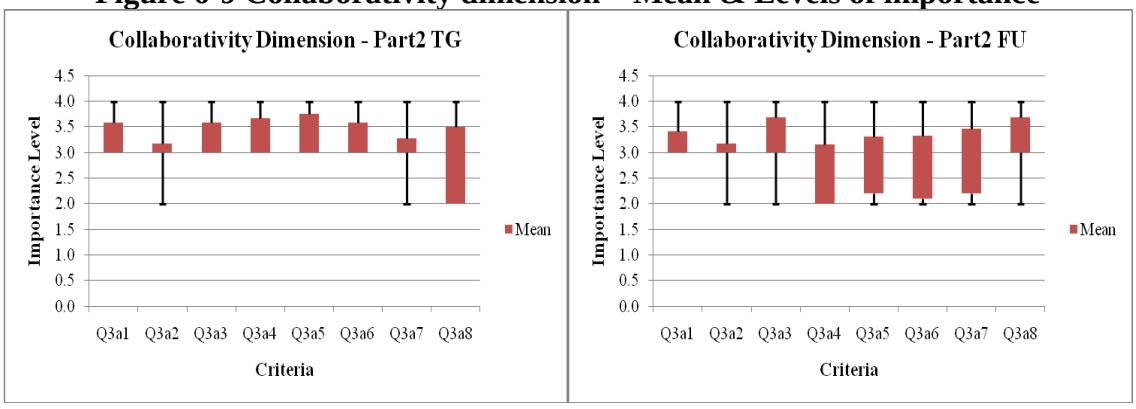
The Collaborativity dimension consists of eight criteria. Table 6-15 depicts that all criteria were ranked at either *very important* or *important* by over 50% of participants in a group. Nobody selected *N/A* and *Not Important*. The two groups' average importance levels are similar, i.e. at *important - near to very important*. TG's range is between 3.17 (for Q3a2) and 3.75 (for Q3a5) whereas FU's is between 3.15 (for Q3a4) and 3.69 (for Q3a3 & Q3a8) (see Figure 6-9).

Table 6-15 Collaborativity Dimension– ranking criteria

Frequency Percentages/Counts Comparison between TG and FU

Level of importance		Not Applicable (N/A)		No Response (N/R)		Not important 1		A little important 2		Important 3		Very important 4	
%/#		%/#		%/#		%/#		%/#		%/#		%/#	
Groups		TG	FU	TG	FU	TG	FU	TG	FU	TG	FU	TG	FU
Criteria	Q3a1	0/0	8/1	0/0	0/0	0/0	0/0	0/0	0/0	42/5	54/7	58/7	38/5
	Q3a2	0/0	8/1	0/0	0/0	0/0	0/0	8/1	8/1	67/8	62/8	25/3	23/3
	Q3a3	0/0	0/0	0/0	0/0	0/0	0/0	0/0	8/1	42/5	15/2	58/7	77/10
	Q3a4	0/0	0/0	0/0	0/0	0/0	0/0	0/0	23/3	33/4	38/5	67/8	38/5
	Q3a5	0/0	0/0	0/0	0/0	0/0	0/0	0/0	15/2	25/3	38/5	75/9	46/6
	Q3a6	0/0	0/0	0/0	8/1	0/0	0/0	0/0	15/2	42/5	31/4	58/7	46/6
	Q3a7	0/0	0/0	8/1	0/0	0/0	0/0	8/1	15/2	50/6	23/3	33/4	62/8
	Q3a8	0/0	0/0	17/2	0/0	0/0	0/0	17/2	8/1	8/1	15/2	58/7	77/10

Figure 6-9 Collaborativity dimension – Mean & Levels of importance



6.3.4 Error Tolerance Dimension

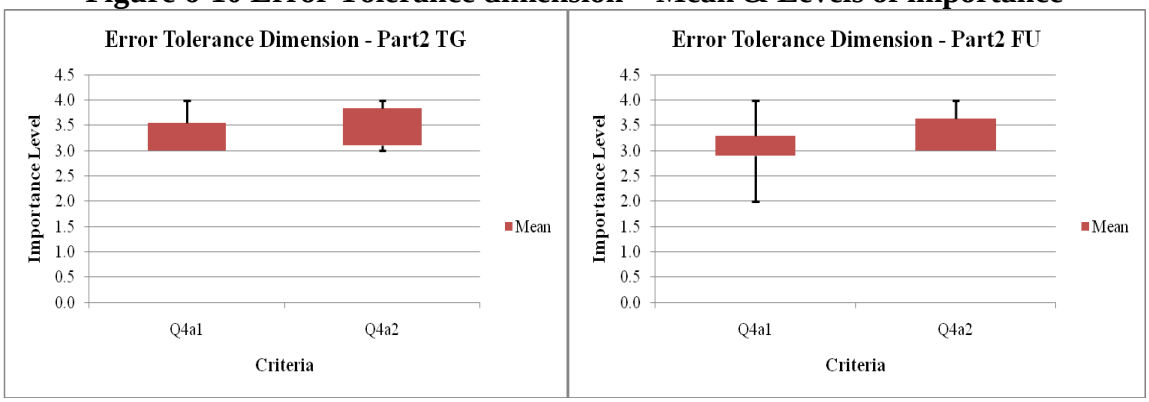
The Error Tolerance dimension has two criteria. Table 6-16 shows that Q4a2 was ranked at *very important* by over 50% of participants in a group while Q4a1 was ranked at *very important* by 6 participants (50%) in TG. Each criterion was ranked at *N/A* by two participants in FU. The two groups' average importance levels are very similar, between *important* and *near to very important*. TG's range is between 3.55 (for Q4a1) and 3.83 (for Q4a2) while FU's is between 3.30 (for Q4a1) and 3.64 (for Q4a2) (see Figure 6-10).

Table 6-16 Error Tolerance Dimension– ranking criteria

Frequency Percentages/Counts Comparison between TG and FU

Level of importance		Not Applicable (N/A)		No Response (N/R)		Not important 1		A little important 2		Important 3		Very important 4	
%/#		%/#		%/#		%/#		%/#		%/#		%/#	
Groups		TG	FU	TG	FU	TG	FU	TG	FU	TG	FU	TG	FU
Criteria	Q4a1	0/0	15/2	8/1	8/1	0/0	0/0	0/0	8/1	42/5	38/5	50/6	31/4
	Q4a2	0/0	15/2	0/0	0/0	0/0	0/0	0/0	0/0	17/2	31/4	83/10	54/7

Figure 6-10 Error Tolerance dimension – Mean & Levels of importance



6.3.5 Universal Accessibility (Ubiquity) Dimension

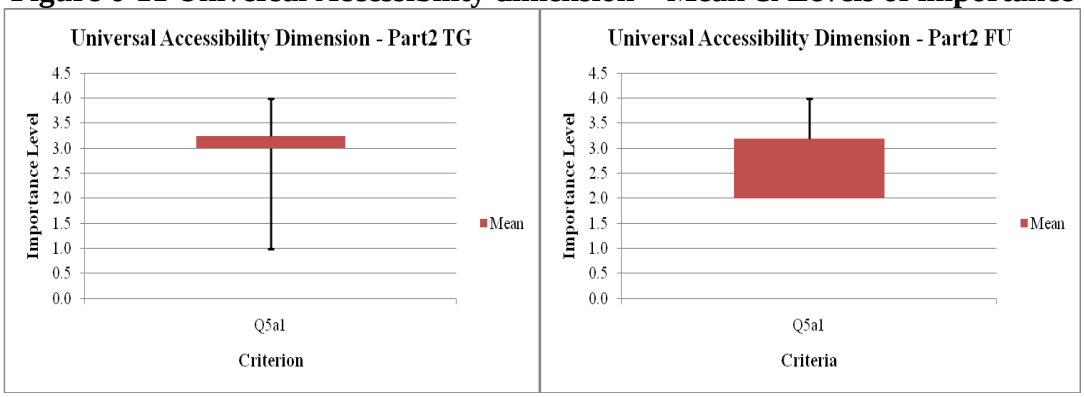
This dimension has one criterion. Table 6-17 shows that Q5a1 was ranked at *important* by 6 participants (50%) in TG. But in FU, the responses spread to different levels. All participants except three participants in FU responded to the question. Nobody selected *N/A* but only one person in TG selected *Not important*. The two groups' average importance levels are very similar, at *important* level. TG's average is 3.25 while FU's is 3.20 (see Figure 6-11).

Table 6-17 Universal Accessibility Dimension– ranking criteria

Frequency Percentages/Counts Comparison between TG and FU

Level of importance	Not Applicable (N/A)		No Response (N/R)		Not important 1		A little important 2		Important 3		Very important 4	
	%/#		%/#		%/#		%/#		%/#		%/#	
Groups	TG	FU	TG	FU	TG	FU	TG	FU	TG	FU	TG	FU
Q5a1	0/0	0/0	0/0	23/3	8/1	0/0	0/0	23/3	50/6	15/2	42/5	38/5

Figure 6-11 Universal Accessibility dimension – Mean & Levels of importance



6.3.6 Satisfaction Dimension

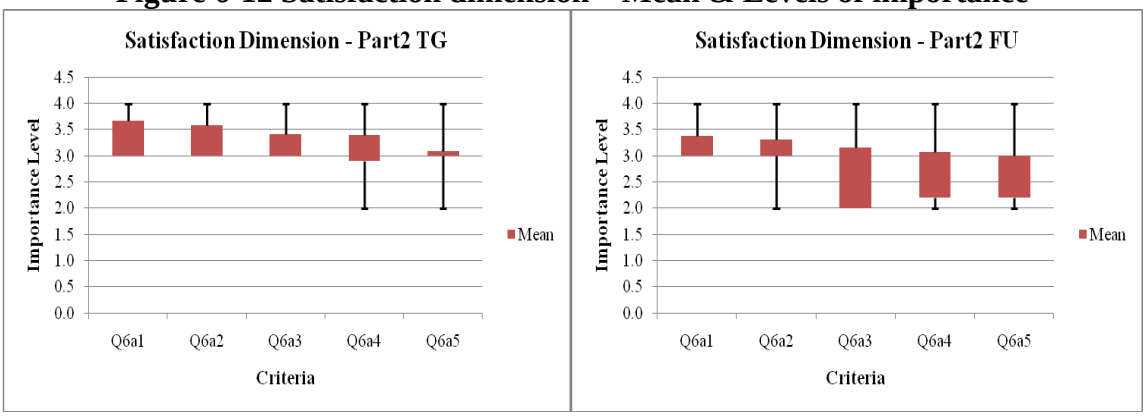
The Satisfaction dimension has five criteria. Table 6-18 illustrates that all criteria were ranked at either *very important* or *important* by over 50% of participants in a group. All participants except two persons in TG responded to the questions and nobody selected *N/A* and *Not important*. The two groups' average importance levels are similar, at *important level*. TG's range is between 3.08 (for Q6a5) and 3.67 (for Q6a1) whereas FU's is between 3.00 (for Q6a5) and 3.38 (for Q6a1) (see Figure 6-12).

Table 6-18 Satisfaction Dimension– ranking criteria

Frequency Percentages/Counts Comparison between TG and FU

Level of importance		Not Applicable (N/A)		No Response (N/R)		Not important 1		A little important 2		Important 3		Very important 4	
		%/#		%/#		%/#		%/#		%/#		%/#	
Groups		TG	FU	TG	FU	TG	FU	TG	FU	TG	FU	TG	FU
Criteria	Q6a1	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	33/4	62/8	67/8	38/5
	Q6a2	0/0	0/0	0/0	0/0	0/0	0/0	0/0	8/1	42/5	54/7	58/7	38/5
	Q6a3	0/0	0/0	0/0	0/0	0/0	0/0	0/0	23/3	58/7	38/5	42/5	38/5
	Q6a4	0/0	0/0	17/2	0/0	0/0	0/0	8/1	15/2	33/4	62/8	42/5	23/3
	Q6a5	0/0	0/0	0/0	0/0	0/0	0/0	8/1	15/2	75/9	69/9	17/2	15/2

Figure 6-12 Satisfaction dimension – Mean & Levels of importance



6.3.7 Summary

Table 6-19 The criteria ranked by 50% or more participants in a group

Ranking levels			Important		very important		very important 4 + important 3	
% (#)			% (#)		% (#)		(% for level 4 + % for level 3) >50%	
Dimen-ions		**Groups**	**TG**	**FU**	**TG**	**FU**	**TG**	**FU**
	Q #	**UE Criteria**						
Effective -ness:	1a1	Completeness			67 (8)	54 (7)	✓	✓
	1a2	Visibility	58 (7)	77 (10)			✓	✓
	1a3	Organisation/Design	50 (6)			54 (7)	✓	✓
	1a4	Navigability	58 (7)			62 (8)	✓	✓
Efficiency	2a1	Speed				54 (7)	✓	✓
	2a2	Familiarity/Consistency/ Standards			67 (8)	54 (7)	✓	✓
	2a3	Flexibility/Adaptability/Configurability	50 (6)	69 (9)	50 (6)		✓	✓
	2a4	Effort	50 (6)	85 (11)			✓	✓
Collaborativity	3a1	User Management		54 (7)	58 (7)		✓	✓
	3a2	Awareness	67 (8)	62 (8)			✓	✓
	3a3	Communication			58 (7)	77 (10)	✓	✓
	3a4	User Control/Moderator & Teacher Control			67 (8)		✓	✓
	3a5	File/Content Sharing/Management			75 (9)		✓	✓
	3a6	Process Tracking/Automated Notification			58 (7)		✓	✓
	3a7	File/Content Protection	50 (6)			62 (8)	✓	✓
	3a8	Security			58 (7)	77 (10)	✓	✓
Error Tolerance	4a1	Error Rate			50 (6)		✓	✓
	4a2	Error Prevention			83 (10)	54 (7)	✓	✓
Universal Accessi-bility	5a1	Support different users	50 (6)				✓	✓
Satisfaction	6a1	Usefulness/Functionality		62(8)	67 (8)		✓	✓
	6a2	Learnability/Predictability/ Recognition/Memorability		54 (7)	58 (7)		✓	✓
	6a3	Simplicity	58 (7)				✓	✓
	6a4	Help/Documentation		62 (8)			✓	✓
	6a5	Aesthetic Design	75 (9)	69 (9)			✓	✓
Total #	**24**		**10**	**9**	**13**	**9**	**24**	**24**

Note: ✓**:** means in a group, (% of the participants who ranked a criterion at "important" level + % of the rest participants who ranked the criterion at "very important" level) > 50%.
#: number of the participants in a group who ranked a level of importance of a criterion
%: percentage of the participants in a group who ranked a level of importance of a criterion
The criteria ranked at "very important" level by over 50% of each group are highlighted in light Aqua colour;
The criteria ranked at "important" level by over 50% of each group are highlighted in light Purple colour.

Table 6-20 The criteria ranked at "Not Applicable" (N/A)

	Groups	**TG**	**FU**
		% (#)	% (#)
Q#	**Criteria**		
Collaborativity dimension			
Q3a1	User Management	-	8% (1)
Q3a2	Awareness	-	8% (1)
Error tolerance dimension			
Q4a1	Error Rate	-	15% (2)
Q4a2	Error Prevention	-	15% (2)

Table 6-19 last page and Table 6-20 summarise the data from Tables 6-13 - 6-18. Table 6-19 shows that both TG and FU had very similar ideas on how important all criteria would be to CSCL system usability. More than 50% participants ranked the 24 criteria either at important or very important levels. The criteria which were ranked at "*very important"* level by over 50% of each group and highlighted in light Aqua colour are: Completeness (Effectiveness dimension), Familiarity/Consistency/Standards (Efficiency dimension), Communication and Security (Collaborativity dimension), Error Prevention (Error Tolerance dimension). The criteria which were ranked at "*important"* level by over 50% of each group and highlighted in light Purple colour are: Visibility (Effectiveness dimension), Flexibility/ Adaptability/ Configurability and Effort (Efficiency dimension), Awareness (Collaborativity dimension), and Aesthetic Design (Satisfaction dimension).

Table 6-20 lists four criteria (Q3a1, Q3a2, Q4a1, and Q4a2) which were ranked as N/A by one or two participants in FU. For all six dimensions, TG's Mean ranges are very similar to FU's. Both TG and FU's ranges are between *important* and *near to very important*.

6.4 Post-test Questionnaire – Part 3, Open ended questions

There are two questions in Part 3 - Qb and Qc as shown below:

Qb: Why do you think the criteria of each dimension are important or not important to a CSCL system's usability (in general)?

Qc: Please write down your comments about each Dimension of OJS usability:

(e.g. what make easy or hard to completing a task/team task? what do you like or dislike? What need to be improved on this system?)

The findings drawn from the responses to Qb & Qc are explained in the order of the six dimensions. The findings directly quoted from the participants' answers are shown in italic.

6.4.1 Importance of the criteria to CSCL system usability –Qb (why?)

6.4.1.1 Effectiveness Dimension

38.46% of FU (5 participants) did not respond to this question.

Key reasons from TG:

Some participants believed that "*Organisation & Navigability – makes users use system more comfortable"; "Each criterion above impacts on the other, e.g. organisation/design affects visibility & navigability, in turn affecting completeness*". So, the criteria will affect time of completing a task. If they are well designed, then a system would be user friendly, usable, comfortable, accessible, and attractive to user. UI is important.

Key reasons from FU:

Some participants thought "*the criteria above are important especially for new users. For instance, without proper navigation button and the visibility of the interface, user will consider it a difficult system, even though it might be simple to use"; "The simple interface, clear visibility with simple navigation make it easy to complete a team task*". Therefore, the criteria will affect the time of completeness, ease of use, trust building, and user experience.

6.4.1.2 Efficiency Dimension

8.33% of TG (1 participant) and 38.46% of FU (5 participants) did not respond to this question.

Key reasons from TG*:*

The criteria will affect ease of use, helpfulness, and reliability of a system. Users will be able to use it, complete a task, and get outcomes quickly, no need for outside support if a system meets the criteria/requirements. Some key reasons noted by some participants are *"Speed is important for users to feel the system responsive"; "With low speed, it is likely that user will not be willing to use the system";* "*Effort is most important; if it takes too much effort, then I would not want to use it"; "If the system can provide a consistency interface then the user can easily get into it"; and "The system should be intuitive enough that outside support should not be needed*".

One participant believed that Familiarity *"is important to a new user but less importance to experienced users"*.

Key reasons from FU:

The criteria will impact completion of tasks in time, how much confusion and frustration users may experience when they work with a system. Users would become familiar with the system's UI quickly if a system meets the criteria. Speed is very important. Some participants commented that "*I feel the speed is most important. The effort is higher than other two because I don't like to get too much limitation or confusion when I am using a system"; "Consistency is also very important, because the system can always tell me where I am. So I can efficiently complete my tasks"; "Consistency of whole system interface is important because it helps user familiar with the system*".

One participant considered Familiarity unimportant because "*as long as navigation is clear, after a while using the system, a user will become familiar with the system anyway*".

6.4.1.3 Collaborativity Dimension

Eleven out of 25 participants (61.54% of FU (8 participants) and 25% of TG (3 participants) did not respond to this question.

Key reasons for TG:

"Collaborativity is important because of working in a group's activity, it is essential"; "Communication and object sharing are key features of collaboration and for it to take place"; "Security is very important to build the trust of the users as it allows the user to feel safe and comfortable when using the system"; "Process Tracking/Automated Notification, will reduce/improve task completion steps". Thus, the criteria are essential to system usability, they are the key features of a system used for group activities. They will affect performance and speed, group cohesion, and collaborative communication. When using a system that meets the criteria, users will feel safe and comfortable.

One participant commented that "*Security and protection are "environmental" issues and not really related to collaboration"*.

Key reasons from FU:

"*The main purpose of such a system should focus on supporting group tasks effectively, so, collaborativity of the system is the most important aspect"; "The user management and communication are most important because I feel the team collaboration is the core part of the system. Easy to touch with the team and members also will make me happy and desire to use the system*". So, the criteria in this dimension are the focuses and the core features of a system supporting group activities; will determine if users can successfully complete team tasks, and if users will be happy to use it.

One participant thought that Security is not an important aspect of a CSCL system as "*While security is important, I don't think that it is necessary for collaborative to take place".*

6.4.1.4 Error Tolerance Dimension

8.33% of TG (1 participant) and 53.85% of FU (7 participants) did not respond to this question.

Key reasons from TG:

People make errors inevitably, therefore, "*handling human errors is a very important aspect in collaborative systems where potentially many members can make error*"; "*If the system provides error alerts, this would make user simply avoid errors*"; "*Error message can stop users doing wrong things and save time and resources*". To sum up, these criteria ensure that a system will help user foresee errors, avoid making mistakes, save time and resources. They will also affect users' willingness to use a system and how easily a system can be accessed.

Key reasons from FU:

"*It is important to make sure that the users continuously work towards the task without any problems*"; "*Prevention is better than to cure*"; "*The criteria are important. If errors are not detected earlier, the system can suffer later which will not be performing as expected. As the saying goes 'prevention is better than to cure', detecting errors and preventing will help users perform and complete their tasks*". So, the criteria will ensure users perform and complete their tasks without problems, help decide what they should do when errors occur.

6.4.1.5 Universal Accessibility Dimension

53.85 % of FU (7 participants) did not respond to this question.

Key reasons from TG:

There are different people with different IT skills in a collaborative environment so "*It is important if the system can be used by a diverse set of users. A good system should be designed for everyone not just for expert users*". "*In collaborative environments, the users should not have to be the experts of the system itself*". The system for teamwork should help different people use the system easily; this criterion would affect whether the system would be popular and cost efficient.

Key reasons from FU:

"*Multiple persons will use the system, each has different characteristics so the system should cater for all.*" "*The system should cover general users; it does not just focus some types of users, because the system should be designed for users who have general computer background*". As a result, users with various IT skills will be able to work on teamwork on the system easily. This criterion will determine whether it is easy to learn.

A participant believed that the criterion is unimportant as it "*is not the main factor of affecting the decision – user acceptance of the system*".

6.4.1.6 Satisfaction Dimension

8.33% of TG (1 participant) and 30.77% of FU (4 participants) did not answer this question.

Key reasons from TG:

These criteria are the core features of a system and the requirements of a system because "*Documentation is very important because without it, it is impossible for a user to learn a system*" and "*Usefulness/Functionality is very important, this is a requirement*". "*Satisfaction is essential to keep the users using the system. If the system produces unsatisfied outcomes like very complicated / and no help from the site for the users, then users will not continuously use that system*". "*If people are satisfied with the products, people will tend to use it more*". The criteria will make sure that a system helps user save time, match their experiences, learn the system, and complete the expected workflow easily. They will also affect users' willingness to use a system.

A participant commented that "*Aesthetic design is not so important as long as the system fits for its purpose".*

Key reasons from FU:

The criteria are the features that a system for teamwork must have as "*they can make sure the system's usability has a high quality*". "*If the system does not functional perform as expected then it is not useful*". "*Simplicity and help are a little more important than others. Simplicity is the basic of the system interface design. I like more functions, however it must be easy to use, clear, and simple*". Consequently, the criteria will affect the quality of a system's usability, whether the system is simple and useful, easy to use, and can be learnt by users. If a system has the features, then users would be satisfied with using the system, and would want to use it.

6.4.1.7 Summary

Table 6-21 The list of numbers of participants did not respond to Qb

Groups	TG	FU
	# (%)	# (%)
Dimensions		
Effectiveness	0 (0)	5 (38.46)
Efficiency	1 (8.33)	5 (38.46)
Collaborativity	3 (25)	8 (61.54)
Error Tolerance	1 (8.33)	7 (53.85)
Universal Accessibility	0 (0)	7 (53.85)
Satisfaction	1 (8.33)	4 (30.77)
Average # (%) of participants (N/R)	1 (8.33)	6.83 (52.56)

Note: #: number of the participants in a group; %: percentage of the participants in a group.

Table 6-21 shows there were more participants in FU than in TG who did not respond to the questions. For Collaborativity, Error Tolerance and Universal Accessibility, more than half of the participants in FU provided no responses on why these three dimensions are important or not important to CSCL system usability.

Table 6-22 next page summarises the reasons why the criteria in the six dimensions are important to system usability. 26 reasons were given by the participants. They can be grouped into two types of reasons: one is related to a system itself and the other is related to users. TG gave more reasons than FU did. Some reasons given by both TG and FU are highlighted in Turquoise colour. The three reasons given by both groups for the importance of the Efficiency dimension are *Time spending on a task*, *Familiarity*, and *No much effort*; the reasons for the importance of the Universal Accessibility dimension are *Completing a task*, *Ease of use and learning*, and *Support various users*; the reasons for the importance of the Satisfaction dimension are *Ease of use & learning*, *simplicity*, and *Willing to use*.

Table 6-22 also shows that the two reasons for the importance of five dimensions, given by TG, are *Completing a task* and *Time spending/speed*. The reason for the importance of four dimensions, given by TG, is *Ease of use & learning* while the reason given by FU is *User experience*. The reason for the importance of three dimensions given by TG is *Core/key/essential features* whereas the reason given by FU is *Ease of use & learning*.

Table 6-22 The list of reasons for the importance of the six dimensions

Dimensions	Effective-ness		Efficiency		Collabora-tivity		Error Toleranc e		Universal Accessibility		Satisfaction		# of dimen-sions	
Groups	TG	FU	TG	FU	TG	FU	TG	FU	TG	FU	TG	FU	TG	FU
Reasons related to a system itself														
Be usable	✓		✓									✓	2	1
Completing a task	✓			✓	✓		✓		✓	✓	✓		**5**	2
User friendly	✓			✓									1	1
Understandable/ logical	✓												1	0
Time spending/ speed	✓	✓	✓	✓			✓		✓		✓		**5**	2
Familiarity		✓	✓	✓									1	2
Ease of use & learning		✓	✓				✓		✓	✓	✓	✓	**4**	**3**
Simplicity		✓							✓		✓	✓	2	2
Reliability			✓										1	
Responsivity			✓		✓								2	0
Providing a lot of help			✓										1	0
Core/key/ essential features					✓	✓	✓				✓		**3**	1
Communication					✓								1	0
Task performance					✓		✓						2	0
Group work						✓								1
Avoid mistakes							✓	✓					1	1
Support various users									✓	✓			1	1
Cost efficiency									✓		✓		2	0
Quality of a system												✓	0	1
Subtotal # of reasons	5	4	7	4	5	2	6	1	6	3	6	4	-	-
Reasons related to users														
Comfortable to use	✓												1	0
Trust building		✓			✓								1	1
Willing to use			✓								✓	✓	2	1
Acceptance				✓									0	1
No much effort			✓	✓									1	1
Knowing what to do next								✓						1
User experience		✓		✓		✓			✓			✓	1	**4**
Subtotal # of reasons	1	2	2	3	1	1	0	1	1	0	1	2	-	-
Total # of reasons	6	6	9	7	6	3	6	2	7	3	7	6	-	-

Note: a reason for a dimension selected by both groups is highlighted in Turquoise colour.

6.4.2 Users' comments about the dimensions of OJS Usability– Qc

6.4.2.1 Effectiveness Dimension

30.77% of FU (4 participants) did not respond to this question.

Key comments from TG:

OJS is simple and efficient, easy to use in terms of navigation and completing teamwork of peer review; email is useful to group communication. One participant commented that it is "*very easy to assign reviewers, and easy to review their comments"* while another person stated that "s*ystem was easy to navigate once I got accustomed it"*.

However improvement on visibility, readability, and navigability, such as hyperlinks/icons/menus, layout structure, pop up screen, etc are required because *"some image links such as 'send mail' and 'review the comments' are hard to be identified by the first look on those images"; "The hardest thing is that OJS is all text-based with only one or two icons representing actions. The border frame – less layout also makes it harder to navigate";* and *"The pop up screen with a reviewer's comments was confusing – I was not sure if I was reading reviewer 1's comments or reviewer 2's".* Some participants suggested that menu should be more visible, contents on web pages should be more readable and colourful.

Key comments from FU:

OJS is an easy and effective system. Users can effectively navigate and complete a task. One participant noted that "*I like the system. It is easy to follow instruction to complete a task".* Another participant also thought that it was "*easy for completing a task"*.

However navigation, layout, menus, and information flow were confusing and not user friendly; participants had to take time to understand and find out how to get started because "*OJS gives a lot of confusions when understanding the processes that need to be done next"; "Navigation system is not user friendly"; "There is a navigation bar at the top and one at the side column, can be a bit confusing"* and *"hard to understand the flow most of the time".*

6.4.2.2 Efficiency Dimension

25% of TG (3 participants) and 46.15% of FU (6 participants) did not answer this question.

Key comments from TG:

"*It was easy to understand and work through the system".* Participants completed the task efficiently once they had learnt and been familiar with the system. But OJS is *"difficult to find some buttons e.g. Log out. The link is located to the right edge of the page. Quite often it is located on the top or the end of a webpage".* So, the links/icons and the structure of the system need to be improved.

Key comments from FU:

OJS is a good, easy, and simple system. It is efficient for users to complete a task; it has friendly interface, provides email templates so users don't need to take a lot of effort to complete a task and save their time. One participant explained that "*what I like about OJS is that it takes less effort to complete a task. For instance unloading and submitting articles was easy. The response time was excellent and navigating through the steps for submitting and reviewing articles is a simple task".* Another participant commented that *"I really like the templates and that save me time".*

However some participants commented that it was hard to understand at a glance, needed help/assistance. Sitemap should be provided and some headings should be added onto the top of each webpage so "*user can easily know structure of web site".*

6.4.2.3 Collaborativity Dimension

33.33% of TG (4 participants) and 38.46% of FU (5 participants) did not give answers to this question.

Key comments from TG:

There are some good features in OJS, such as showing the progress status like articles in review, if a user has been assigned or unassigned to a role, and the date stamp when an action is completed. One participant commented that "*after logging in I could see how many articles in review, unassigned, etc".* Another participant stated that *"Roles and the actions of other users are visible and can be traced".*

However, "*the workflow is not well presented. File upload dialogs can appear in more than one place of a page which is a bit confusing".* Users got lost easily. Security and awareness need to be improved.

Key comments from FU:

Users can easily complete a task collaboratively, such as communicating with other teammates by email, making comments, tracking a process, and managing files. One participant noted that "*communicating with other team mates is easy. Reviewing other group works and making comments on their work are easy. The system enables user to be able to download and upload files; sending message to other group mates is great".* But the improvement on OJS should focus on increasing communication channels and traceability because "*in this system, I only see one communication channel, that is email. However there are lots of channels can be joined such as instant message, real time chatting, video and sound". And "individual user uploaded files or contributions to the group are not traceable".*

6.4.2.4 Error Tolerance Dimension

33.33% of TG (4 participants) and 53.85% of FU (7 participants) did not answer this question.

Key comments from TG:

There are some instructions how to withdraw the wrong actions and irreversible actions that are followed by a 'confirm' dialog. One participant commented that "*I think the system was pretty good at allowing me to withdraw from the errors I made, which were mostly navigation errors".* However, some participants thought that OJS had no warnings and error alerts when they made mistakes. For example, I "*did not see this feature. There seemed to be no warnings".*

Key comments from FU:

OJS provides the information about what should be done to avoid a mistake. However Negative comments indicated OJS had poor error tolerance. Adding reverse functions and increasing the capability of this dimension on OJS were suggested. e.g. "*Error message should be displayed in an obvious location when a mistake is made" and "I feel the "undo" "cancel", "reverse" is good for users".*

6.4.2.5 Universal Accessibility Dimension

33.33% of TG (4 participants) and 53.85% of FU (7 participants) did not answer this question.

Key comments from TG:

There are only negative comments on this dimension of OJS, i.e. some universal navigational elements like links and buttons were not there, the designs of some links or

menus were not consistent. One participant criticised that "*some icons were hard to be identified. Universal accessibility is not available".* One participant suggested that OJS *"Needs more consistency – some buttons are present but most of the time I found I was clicking on links".*

Key comments from FU:

"The universal accessibility in general is ok. What makes the system easy in this case is that a user is able to quickly know what task is required by going to the editor section and he/she should be able to decide either to review or submit an article". So, OJS is easy regarding these features of the dimension. However, users were unable to configure the system based on their needs. "*The system did not clearly display options so that each user could not configure the system to their needs. This would be a good new feature if it is there".* A participant commented there is no need for improving this dimension on OJS.

6.4.2.6 Satisfaction Dimension

16.67% of TG (2 participants) and 30.77 % of FU (4 participants) did not answer this question.

Key comments from TG:

Some participants were satisfied because they could use it to fulfil tasks easily with little help, e.g. "*I was satisfied with the system usability it was fairly simple to use".*

But there were more negative comments than positive comments in TG. Some participants noted that OJS did not provide users with group email, a simple system structure, integrated help and tips, and a strong search engine. Some participants were confused with a few features of user interface and believed that page layout and design, and text-based web pages were not user friendly. Here are some relevant comments: "*The page layout does not look very friendly, and it should have allowed an option for sending a group email to the reviewers"; "The UI is the main downside of OJS. It does not indicate (well enough) the steps needed to perform a task. ... The layout and design is a bit confusing. Lots of text needed to go through to find the right commands".*

Key comments from FU:

There were more positive comments than negative comments in FU. Users were able to use it. The system was easy and simple to understand, learn, and use for completing collaborative tasks, especially for submitting and reviewing articles. It was a useful

environment for online journal review and had a great and innovative way for users to work in a team. Email tools, such as email templates and a hyperlink that links to a file to be reviewed and shown in an email, saved users' time when completing a team task. One participant commented that "*From my personal point of view the overall satisfaction of the system is great. Using OJS to work on an assignment in a collaborative way is easy. The system functions as expected. I was satisfied when I was able to successfully submit an article and review an article".* Another participant advised that *"I like the templates that save me time. Some link buttons such as 'email link' is good".*

Some participants suggested that the improvement on flexibility, colour design and documentation should be carried out in future. For example, "*OJS needs more flexibility, and better colours".*

6.4.2.7 Summary

Table 6-23 The list of Numbers of participants did not respond (N/R) to Qc

Groups	**TG**		**FU**	
	# (%) (gave N/C / N/A)	# (%) (N/R)	# (%) (gave N/C / N/A)	# (%) (N/R)
Dimensions				
Effectiveness	0	0	0	4 (30.77)
Efficiency	0	3 (25)	0	6 (46.15)
Collaborativity	1 (N/A)	4 (33.33)	0	5 (38.46)
Error Tolerance	0	4 (33.33)	0	7 (53.85)
Universal Accessibility	1 (N/C)	4 (33.33)	0	7 (53.85)
Satisfaction	-	2 (16.67)	0	4 (30.77)
Average of # (%) participants (N/R)	-	2.83 (23.58)	-	5.5 (42.31)

Note: **N/A**: Not applicable; **N/C**: No comments; N/R: No response.
#: number of the participants; %: percentage of the participants.

Table 6-23 shows that there are more participants in FU than TG who did not respond to this question. On average, there are 42.31% of FU (around 6 participants) and 23.58% of TG (around 3 participants) who did not provide their comments on the six dimensions reflecting the features of OJS.

Table 6-24 below lists the types of comments on the features of each dimension of OJS. Regarding satisfaction dimension, the participants in FU gave positive comments on most criteria but the participants in TG were not happy with this dimension. Both groups gave negative comments for the following criteria: Organisation/design, Navigability, Security, Error Prevention, Support different users,

Usefulness/Functionality, and Help/Documentation but provided positive comments for Ease of use, Navigability, Email tools, File/Content Sharing/Management, Error Prevention, Usefulness/Functionality, and Simplicity. "Ease of use" was commented as a good feature of four dimensions of OJS.

Table 6-24 comments on the six dimensions of OJS usability from TG and FU

	Groups	**TG**		**FU**	
	Comments	**Positive** ✓	**Negative** ×	**Positive** ✓	**Negative** ×
Dimensions					
Effectiveness	Completeness				
	Visibility		×		
	Organisation/Design		×	✓	×
	Navigability	✓	×	✓	×
	Helpfulness	✓			×
	Ease of use	✓		✓	
	Usefulness	✓			
	simplicity	✓			×
	Email / communication	✓			
Efficiency	Speed			✓	
	Familiarity/Consistency/Standards		×		
	Flexibility/Adaptability/Configurability				
	Effort		×	✓	
	Ease of use	✓		✓	
	Helpfulness	✓			×
	Understandable	✓			×
	Tools / templates			✓	
Collaborativity	User Management	✓			
	Awareness		×		
	Communication/email	✓		✓	×
	User Control/Moderator & Teacher Control				
	File/Content Sharing/Management	✓		✓	×
	Process Tracking/Automated Notification				
	File/Content Protection				
	Security		×		×
	Ease of use			✓	
Error Tolerance	Error Prevention	✓	×	✓	×
Universal Accessi-bility	Support different users		×	✓	×
	Ease of use			✓	×
Satisfaction	Usefulness/Functionality	✓	×	✓	×
	Learnability/Predictability/ Recognition/Memorability		×	✓	
	Simplicity	✓		✓	
	Help/Documentation		×		×
	Aesthetic Design			✓	×
	Navigation		×		
	Email templates			✓	

Note: The features in italic were added to the dimensions according to participants' comments; A comment given by both TG and FU is highlighted in turquoise colour.

Both groups suggested that the improvements on the features determined by the following criteria should be done:

- Navigability, User interface design, and Visibility in Effectiveness dimension;,
- Consistency and Familiarity in Efficiency dimension;
- Security, Awareness, Communication, and Traceability in Collaborativity dimension;
- Error prevention in Error Tolerance dimension;
- Universal design of links and menu in Universal Accessibility dimension;
- Functionality, Flexibility, Recognition, and Help/documentation in Satisfaction dimension.

6.5 Task sheet and Observer data collection sheet

6.5.1 Findings

The TS and ODCS are shown on Appendix 16 and Appendix 17. A TS was filled in by each participant in TG after he/she completed the testing task while an ODCS was completed by the facilitator/observer when a participant was working on the testing task and was observed. The participants did not answer the question 7 in ODCS because it was similar to Qc in Part 3 of Post-test Questionnaire and the question 6 for Interview.

As explained in Section 4.8 Step1 on pg 64, the findings were drawn from the raw data that was recorded in TS and ODCS and the digital audio and video files (screen capture files). These findings are represented in the sequence of the questions on both TS and ODCS in this section. The findings directly quoted from the participants' answers are shown in italic.

6.5.1.1 Completion of the testing task

TS -Q1: Have you completed the task successfully? TS-Q2 Is this task hard or easy?
Eleven participants (91.67%) completed the task during their testing sessions. 10 participants (83.33%) thought that the testing task was easy. One participant (8.33%) was undecided if it was easy or hard.

ODCS-Q1: Has the task been completed?
Eleven participants (91.67%) completed the task either at round 1 or at round 2 of the review. Only one participant (8.33%) did not complete it.

6.5.1.2 Tools availability on OJS

TS-Q3: Does the system provide easy ways or tools to help you complete the task? The ways / tools are:

Nine participants (75%) agreed that OJS system provided some easy ways or tools to help them complete the task, and also identified what the tools were, such as *"Easy navigation, email & attaching the reviewer's comments and a link to a file to be reviewed with an email, showing a status of a review process".* One participant (8.33%) undecided if the tools were helpful or not while two participants (16.67%) disagreed that they were helpful. One participant criticised that *"the tools are all text-based which makes it hard to find some commands. A graphical representation of the review process would help"* and the other commented that "*the system is not final yet and awful".*

6.5.1.3 Mistakes made during a testing

TS-Q4: Did you make mistake when doing the task? The mistakes are:

Only one participant (8.33%) thought he did not make any mistakes and the rest participants admitted that they made a few mistakes when completing the testing task. The common mistakes were related to navigation, emailing, uploading files, assigning roles, missing steps, etc.

ODCS-Q4: What the mistakes are? How much time spent on correcting them?

Most mistakes found during the testing sessions were either clicking wrong links/icons or missing one or more steps shown in the task script. For example, seven participants (58.33%) clicked wrong links or buttons when uploading a file and sending an email. Four participants (33.33%) skipped a step - inserting reviewers' comments in an email, so they had to go back and do it again.

However most participants spent a short time (from a few seconds to several minutes) on finding the ways to correct the mistakes, e.g. spending a few seconds to find the right icon for sending an email, spending a couple of minutes to upload a file, etc.

6.5.1.4 Teamwork involved in the task

TS-Q5: was this task involved team work?

One participant (8.33%) said *not*. The rest of participants noted down "*Yes*".

TS-Q5a: If yes, how did you communicate with your teammate via the system? (e.g. email)

10 participants (83.33%) thought that Email was the communication tool in OJS.

TS-Q5b: Does the system help you collaboratively work with your teammates? Why do you say so:

12 participants (100%) agreed that the system helped them work collaboratively with their teammates as they *"were able to communicate among team members, to assign roles and allocate tasks, to review and comment on others' work, and to receive teammates' comments and suggestions by email".*

ODCS-Q2: Does the task involve teamwork? Q3 how many users are there in the team?

Yes. If the process ended at round 2 of peer review, then 4 teammates would be involved in a team. If the process ended at round 1, then 3 teammates would work together in a team.

6.5.1.5 Questions asked during a testing

ODCS-Q5: What questions did the user ask when he/she was working on the task?

Participants asked different questions during their testing sessions. These questions can be put into two main groups.

1) The questions related to using OJS: some common questions were "*where is the mail icon? Log out button? Resubmit button? Can I send an email to two reviewers at a time? where to add myself as an Editor?*".
2) The questions related to the process of peer review: e.g. "*how to make a decision? how can I get to Round 2? What am I supposed to do after I select the decision 'resubmit for review?*"

6.5.1.6 Problems/issues

ODCS-Q6: Problems / issues encountered during the testing sessions

The following problems or issues encountered during the testing session:

1) OJS lacked some features supporting navigation and notifying information.
   For instance, when "Assign a reviewer" was clicked, the screen moved up to the top of the web page and did not go back to the section where the participants worked on; when clicking the Back button on the Internet Explorer's menu bar, the screen did not go back to where it was and all the data entered before clicking disappeared.

One participant complained that "*if I clicked, something failed; it should come back where it was but it did not".* Another example, it did not support opening multiple windows at a time**,** so the participants were unable to read the comments from two reviewers side by side. One participant suggested that *"it would be a good feature of collaboration if it is able for users to see both reviewers' comments at the same time".* Moreover, once a file had been uploaded, OJS did not notify users if uploading had been completed or not. One participant commented that "*I would like to see something saying 'completed successfully"*".

2) OJS had poor visual design.

There were many hyperlinks and various font sizes on a page; clickable text and buttons were not very obvious and users got confused easily. The participants did not realise that some texts were clickable until they spent some time on checking. A typical example was 'Upload' button and 'Browse' box. A participant should click 'Browse' box first, select a file and then click Upload button to upload a file. If the user just clicked Upload button, nothing would happen. One participant commented that 'Browse' box "*looks like an editing box and does not look like clickable".*

3) Most participants were unfamiliar with OJS.

Most participants in TG had not accessed OJS before they came to their testing sessions. They had to spend some time to find out where to go, e.g. some could not find "Log out" button, and the link "add self as the Editor". So they had to scroll up and down many times on the screen until they found the right buttons or links.

4) Most participants did not know much about the process of peer review.

The participants knew little about how to do peer review and make a decision on publishing. They had to rely on the instructions shown on the task script.

5) Some participants had improper working manners.

A few participants were reluctant to read instructions and the facilitator had to repeatedly remind them to check the instructions. Due to their reluctance to follow the steps and instructions, some participants made many mistakes and had to spend more time on completing the task. One participant explained that *"I would like to take immediate action without reading the instructions and remembering what should be done".* One did not concentrate on the testing and this resulted in either getting lost or missing steps many times during the testing session. In contrast, two participants read the instructions carefully; they did not make many mistakes, and spent 16mins and 22 minutes respectively to complete 2 rounds of peer review.

6.5.2 Summary

Eleven participants (91.67%) completed the testing task during their testing sessions and 10 participants (83.33%) considered it an easy task. Nine participants (75%) agreed that OJS system provided some easy ways or tools to help them complete the task.

Eleven participants (91.67%) agreed that the task involved team work. 10 participants (83.33%) considered Email as the tool for communication with teammates and 12 participants (100%) commented that the system helped them work collaboratively with their teammates in OJS.

The questions often asked by participants were related to either using OJS or the process of a peer review. There are two common types of mistakes: clicking wrong links/icons and missing steps. The problems encountered during the testing sessions are OJS lacked of some features in supporting navigation and notifying information and had poor visual design, most participants were unfamiliar with OJS and the peer review process, and some of them had improper working manners.

6.6 Interview

The interview questions are shown in Appendix 5. Twelve participants in TG were interviewed after they completed the testing task. The findings are explained in the order of the following three aspects: OJS CL usability, the framework and criteria, and the user testing. The findings directly quoted from the participants' comments are shown in italic.

6.6.1 OJS CL usability

6.6.1.1 Comments on OJS usability

Q1 What do you think of the usability of the system? What do you like or not like, why?

Regarding OJS usability, seven participants (58.33%) liked OJS because they thought that it was a good system, simple, and easy to use. Three participants (25%) were not quite sure but considered that the system was not too bad. Two participants (16.67%) did not like OJS, particularly its user interface.

The reasons of liking the system are OJS having some good features (six participants (50%)), ease of use (five participants (41.67%)), and good user interface (two

participants (16.67%)), whereas the reasons of disliking the system are poor user interface and navigation (complained by seven participants (58.33%)), lack of error prevention (three participants (25%)), and the text-based web pages (two participants (16.67%)).

6.6.1.2 Comments on collaborative teamwork in OJS

Q2 Regarding to collaborative teamwork, do you think this system is a good system? Why?

Eight participants (66.67%) agreed that OJS was a good system for collaborative teamwork while two participants (16.67%) disagreed, and two participants (16.67%) were not sure and thought it was not too bad.

Some features that were suitable for team work were identified, e.g. email for asynchronous activities was identified by eight participants (66.67%), role administration by 6 participants (50%), storing & sharing files and ideas by 4 participants (33.33%). In contrast, a few participants did not think that OJS was a good system for collaborative teamwork. One participant (8.33%) believed it had poor user interface while 3 participants (25%) thought OJS did not have the feature - error alert and should be built in the features - supporting synchronous activities.

6.6.1.3 Tools/ utilities for collaborative teamwork on OJS

Q3 Do you think the system provides enough information, tools/utilities for supporting collaborative teamwork? E.g. decision making. Why do you say so?

Ten participants (83.33%) agreed that the system provides enough information/tools for supporting collaborative teamwork. One participant commented that "*for this purpose – Journal peer review, it provides enough tools and information."* The reasons for agreeing on "yes" for this question are *OJS*

1) *had a storage where stores submissions, recommendations, decisions, ideas, comments; it would be good and easy for finding information and sharing views/comments in one place and for making decisions;*
2) *gave user the information and steps about how to do things e.g. how to submit an article;*
3) *provided email templates, so users could quickly send out emails to team members and share ideas.*

The reasons for disagreeing on "yes" are: OJS did not have some features which participants would like to have, *"like Wiki, able to edit files online; collaboration that could be made on real time"; "If a user wants to have a discussion about comments and submissions, he/she would want instant messaging".*

Q7: Communication via email only, do you think it is ok for collaborative team work?

Nine participants (75%) confirmed that it was ok for users to communicate with others via email regarding collaborative team work. They thought: *"it is "sufficient for collaborative activities"; "it is ok to the processes of decision making and reviewing paper"; "it (email) is the right option for me"; "Most people communicate via email, email is a very good option. Otherwise everyone have to be in one place in real time"; "it is the feature of an asynchronous system, It would be better to let users get emails first, then know what they are required then logon".*

However, some participants did not agree with the statement: "*email is sufficient for collaborative team work*". They argued that a system for collaborative teamwork should have some other built-in features, such as group email, forum, instant messenger, real time communication, etc.

6.6.1.4 Suggestions for improvements in OJS

Q6 Suggestions about improvement on OJS usability, i.e. a CSCL system usability, in terms of "effectiveness, efficiency, collaborativity, error tolerance, universal accessibility (ubiquity), and satisfaction".

The aspects of OJS CL usability that need to be improved in future are:

1) interface and layout, visibility and navigation in the Effectiveness dimension.

   Some participants suggested that menus and section headings would need better design, and in OJS *"navigation and visibility"* should be improved; *"navigation needs more clarification on where you are";* about "*The email icon*  *and comment's icon*  *, using universal icons, e.g. using an envelope icon, would be more effective because everyone can recognise what they are"; "the layout, font and text need to be improved";* OJS would need *"more visible menus, ... a sitemap page; perhaps a listing of steps under each task"*.

2) consistency and standards and effort in the Efficiency dimension.

   One participant recommended that "*one system, one standard, people don't have to study before using it each time*".

3) some functions supporting real time communication and file versioning in the Collaborativity dimension.

   One participant stated that "*for group work, it definitely needs synchronisation. … like forum, students can question about assignments, talk to one another, post questions, somebody else answer them. This is collaborative learning";* some participant suggested that they *"would like to see more features e.g. instant messages", "have group email set up", "real time interaction", "online chatting", "file version control"*, and " *able to see both reviewers' comments at the same time*".

4) error prevention in the Error tolerance dimension.

   Some participants pointed out that *"Error alerts need to be improved"; "If error occurs, I should get a message to warn me"; "the site should provide some sort of feedback about next steps"*.

5) support different users with different levels of IT expertise in the Universal accessibility dimension:

   "*one system, one standard*" was suggested by TG. One participant recommended that *"Using universal icons e.g. using an envelope icon, would be more effective because everyone can recognise what they are".*

6) recognition and memorability, help and notification in the Satisfactions dimension.

   Some participants suggested to *"make the workflow more intuitive, I don't have to remember what I have done and what I should do";* OJS should *"have a graphic web presentation showing what a user should do, like a timeline";* and have a notification *"saying 'completed successfully once a user has uploaded a file"*.

Q8 would you like to use this system if you have group assignment/work?

Five participants (41.67%) confirmed they would use the system for working on their group assignments and three participants (25%) specified they would use OJS only for reviewing assignments while two participants (16.67%) said they would not like to use the system.

The reasons that some participants would like to use OJS are that OJS was an effective system, and easy to use (five participants (41.67%)), could store & share files and ideas, supported asynchronous communication (email tool & templates), had functions for paper review and role administration (two participants (16.67%) in each group). Some participants commented that "*if we have opportunity to do this sort of the task, I will*

recommend it, because the interface is very clear, friendly"; "I think the system is essential very effective"; "it is simple, ease of use, exchange email and file and everything"; "pretty straightforward, my group basically used email"; "the best thing is clear roles, able to keep documents, send email, can download and upload files, and store comments".

However, the reasons of some participants not wanting to use OJS are that OJS did not support synchronous activities (given by five participants (41.67%)), it needed some features like error alert and forum (given by three participants (25%)), and it had poor user interface design (layout /navigation) (given by one participant (8.33%)). The following comments illustrate why some participants disliked the system: *"Email may be slow"; "it would be only helpful if you do review process"; "you perhaps want something real time, as you want quick feedback".*

6.6.1.5 Satisfaction with OJS

Q9 Overall are you satisfied with OJS?

Eleven participants (91.67%) said they were satisfied with OJS usability. They agreed that it was *"Good for the purpose"* and *"Easy to use",* and liked the email tools.

6.6.2 The framework and criteria

6.6.2.1 Importance of the criteria

Q4 Do you think the criteria shown on the framework are important to UE? What are important or unimportant to the usability of a collaborative learning system?

All participants agreed that the criteria in the framework are important to system usability and UE. The common comments are: "*All criteria are important"; "Effectiveness, Efficiency, Collaborativity are important", "Simplicity is very important";* "*For Collaborative system, all are very important"*. One participant considered *"Error tolerance should be important" but* another participant believed that *"Error tolerance is less important than others".* About the criteria - Security, one participant commented that *"Security is not real important in this case, I think not many people or hacker would like to spent time and money for articles, they would spend time and money on attacking business/bank accounts. We don't need to spend a lot money on improving the security of this system".*

6.6.2.2 The range of the criteria in the framework

Q5 Do you think the framework mostly covers important criteria? Why do you say so? What else do you think needed to be included in the framework?

Except one person, all participants thought the framework "*pretty much cover everything, includes necessary criteria, very details, and it is a good set".* One person suggested adding "backup" as a new criterion but another person thought "*it has too many*". When they were asked what else could be added into the framework, several persons mentioned that they could not think of anything else.

6.6.3 Suggestions on the user testing

Q10 what comments/suggestion on the testing do you have?

Five participants (41.67%) said that it was a good testing. *"I guess it is hard to design a testing for so many different levels of users as you don't know beforehand if a user is very used to the system or not". "I think, the procedure is pretty good, you have had on the right track except the system". "You will get a lot of data from interview and other records".* Several suggestions were collected and listed as below:

1) The time for a testing session should be longer.
   Several participants advised that *"no time to think"," just follow the instructions"*, and *"hard to get the whole picture of the system".*
2) The participants should be given the task script before the testing sessions.
   So, they would not need to read the instructions all the time while working on the testing task, and would have more time to get to know the whole system. Here are some relevant comments given by the participants.
   "Perhaps it is hard to get opinions about the system from following step by step instruction for the first time user, because I concentrated on reading and doing"; "I don't have an abroad opinion as I only worked on the task – decision making. So, I don't have much knowledge of the system as a whole from only the task. it is hard to get the whole overview of the system"; "I would like to get explanation -how to do the task, and then to implement or testing the system"; "possible send email to participants about the system link so they would be able to try it on prior to the testing"; and *"Most of the criteria was not tested in the step by step so I could not have a throughout opinion on them".*
3) The task script should not provide the steps with too specific instructions.
   Some participants explained that "*The instructions are too precise, the instructions did not make me to think myself, I spent more time on looking at the paper";*

"perhaps it is hard to get feedback from testing by following instructions, I don't have a broad view on the system"; "I think having a step-by-step instruction set removed my own thinking on how it worked"; "a first time user won't be able to get whole picture, instructions are too specific, should let user to decide how to do task"; and "because you gave me the instruments – how to do the task, it might cause bias. I am not familiar to the system, I just followed the instructions";

Some participants suggested that *"to design the testing, I would design step by step with more general instructions, instead of say click the exact link, it would test more how people would do/use"; "you should create a help doc and perhaps create work flow before people use it. Perhaps ask different people to follow this flow or that flow/process; should have some screen prints on the instructions, get somebody to read the document / instruction to check if it is readable and understandable. ... You can see the document actually improves the usability".*

4) The testing task should have more users involved.

   It was only two participants involved in the testing task. Some participants recommended the testing should *"have a couple more users to check it out if it would be useful"* and *"have a couple more users to disconnect and connect,* (and) *check things out".*

5) some open-ended questions in the part 3 of Post-test questionnaire should be defined more clearly.

   A few participants pointed out that *"the questionnaire, I was thinking of everything focuses on the system OJS. You can make it clear –for this, you need to answer the question based on the system (OJS) that you are testing, and for that, it is about the kind of systems in general. If you explained more to me, I would have done good feedback on it"; "When I read the question – rank the importance of each criterion of collaborativity to a system's usability, it is usability that is the question. But I only look at collaborative features, how they affect team work. That was my thinking when I started answering the question. I had to read the question again, to see really how it does affect usability".*

6.6.4 Summary

Twelve participants in TG were interviewed after they had completed the testing task. Eleven participants (91.67%) were satisfied with OJS usability and 10 participants (83.33%) thought that it was a good or not a bad system for asynchronous collaborative team activities that do not need real time communication but aim to share information

and ideas, review papers and publish articles, etc. OJS provided the information and the tools needed for completing some tasks, including email and email templates for asynchronous communication, and had the storage for keeping files, comments, and decisions in the system. However, two participants (16.67%) did not like OJS due to its unfamiliar user interface, asynchronous communication, and lack of error prevention. 66.67% (8) participants indicated that they would like to use this system for their group assignment/work in future.

Regarding the framework, all participants agreed that the framework and its criteria were important to UE, the framework covers most important criteria. Another criterion - backup of the system, was suggested to add to the framework by one participant.

Many suggestions about the improvement on the six dimensions of OJS usability were collected from the interviews, such as: changing the icons, links, interface and layout and colour design to improve visibility and navigation and design in the Effectiveness dimension; using one standard for the journal website to increase consistency and enhance familiarity in the Efficiency and Universal accessibility dimensions; adding some features/functions like forum, real time interaction (e.g. instant message & online chatting), and file version control to advance Collaborativity; giving error alert/warning to improve Error prevention; and providing some graphic interface and help instructions/documentation to raise user Satisfaction with OJS.

Regarding the testing, five participants (41.67%) commented that this was good testing. Some participants suggested that they would like to have more time for a testing session, and to receive the task script without including too specific instructions before the testing sessions. They also suggested that it would be better to get more users to work on the testing task during a testing session and to have more clearly definitions on the open-ended questions in the part 3 of Post-test questionnaire.

6.7 Summary

This chapter has presented the findings drawn from the data that was collected from the Pre-test questionnaire, Part 1, Part 2 and Part 3 questions in Post-test questionnaire, the task sheet and the observer data collection sheet, and the interviews.

The majority of participants ranked their CSCL system expertise levels as either level one (novice) or two (beginner). Most participants in TG had not accessed OJS before the testing. The findings in Part 1 identify 16 statements that were either strongly agreed or agreed by more than 50% of the participants from each group. Six statements about Help/Documentation, File/Content Protection, Security, and User Management were ranked as N/A by 50% or more of a group. The average ranges of ranking the agreements on the statements about the criteria in the five dimensions of OJS CL usability (the Error Tolerance dimension excluded) are between *near to undecided* to *agree*. The average ranges of ranking the agreements for the Error Tolerance dimension are between *near to disagreed* to *disagreed* in TG and between *near to undecided* to *undecided* in FU.

The findings from Part 2 prove that all the criteria were either important or very important to the system usability. Both TG and FU groups had similar insights on the importance of the criteria in the framework. For the six dimensions, TG's average ranges of rankings are very similar to those of FU. The ranges of both TG and FU are between *important* and near to *very important*. Part 3 of Post-test questionnaire has two questions – Qb and Qc. Qb is about the reasons of the importance of the criteria to system usability and evaluation. Twenty six reasons were collected and can be categorised into two groups – one is related to the system itself and the other is related to users. The criteria in the six dimensions would determine task completion and the time spent/speed, ease of use and learning**,** user experience, and core/key/essential features of a system. Qc is about the participants' opinions on OJS usability. FU gave more positive comments on the criteria/features in the six dimensions of OJS CL usability than TG did. Both groups gave negative comments to the following criteria: Organisation/design, Navigability, Security, Error Prevention, Support different users, Usefulness /Functionality, and Help/Documentation but provided positive comments to Ease of use, Navigability, Email tools, File/Content Sharing/Management, Error Prevention, Usefulness/Functionality, and Simplicity.

The findings from the responses recorded on TS and ODCS show that 11 participants (91.67%) agreed that this task involved a team work, and 12 participants (100%) thought that OJS helped them work collaboratively with their teammates. Twelve interviews were completed by the participants in TG. Eleven participants (91.67%) were satisfied with OJS usability and 83.33% (10) participants thought that it was a good

system or not a bad system for asynchronous collaborative team activities that do not need real time communication. All participants in TG believed that the framework and its criteria were important to UE. The findings from the interview also provided the suggestions on what should be improved on OJS usability and the user testing. The aspects of OJS CL usability that should be improved are navigation, visibility, interface and layout, consistency and file versioning, real time communication, effort, recognition and memorability, and help/documentation and notifications. About conducting user testing in future, the facilitator was suggested to allocate the proper time for a testing session, to give the participants the task script that do not have too specific instructions before testing, to have more than two users involved in a collaborative task for testing a CL system's usability, and also to ensure that the open-ended questions in post-test questionnaire are clearly defined.

Chapter 7

7. Discussion

The chapter has the following six sections: discussing the three research questions and the framework in detail, an explanation of application of the methodology for this study, and finally a summary of the main points.

7.1 Research question 1

Does the proposed framework consist of important criteria for CSCL system UE in a collaborative teaching and learning environment?

The aim of this question in this study is to find out if the criteria defined in the framework are important to CSCL system UE in an educational CL environment. in this case study, OJS was set up as a CSCL system in SCMS at AUT and used by the students who studied the CC course and accessed the system to work on their group assignments and complete peer reviewing of their assignments. The framework that was tested and modified in the trial testing and pilot study was used to evaluate OJS CL usability by the participants who were mainly recruited from the CC class. The findings presented in Section 6.3 (Part 2 of Post-test questionnaire) and Section 6.4.1 (Part 3-Qb of Post-test questionnaire) and Section 6.6.2 (Interview) will be discussed in this section.

7.1.1 Importance of the criteria to UE

The findings from Q4 in the interviews show that all participants in TG agreed that all criteria are important to a system UE (see Section 6.6.2.1). The findings in Section 6.3.7 illustrate that more than 50% of participants in each group ranked the 24 criteria as either important or very important. The criteria rated as “very important” by over 50% of participants from each group are Completeness (Effectiveness dimension), Familiarity/Consistency/Standards (Efficiency dimension), Communication and Security (Collaborativity dimension), Error prevention (Error Tolerance dimension),

while the criteria rated as 'important' by over half of participants from each group are Visibility (Effectiveness dimension), Flexibility/Adaptability/Configurability and Effort (Efficiency dimension), Awareness (Collaborativity dimension), and Aesthetic Design (Satisfaction dimension) (as shown in Table 6-19 on pg 103).

With regard to whether the framework consists of mostly important criteria for a collaborative system UE, the findings from Q5 asked in the interviews disclose that 91.67% of TG (11 participants) said "*yes*" and agreed that the framework consists of mostly important criteria. However, the answer "yes" or "agree" might not truly reflect their understanding and experience in using the framework to evaluate OJS. The participants in TG had to finish a pre-defined task within a limited time as well as complete questionnaires and an interview, so they might not have sufficient time to access the system and consider the usability evaluation thoroughly. Several participants mentioned that they could not think of any other criteria. Furthermore, some participants might not have got enough knowledge about CSCL system usability and usability evaluation and might not have known what should be included in a framework for UE according to their expertise levels shown in Table 6-4 on pg 88. The answer "yes" or "agree" in such cases might have been a guess or chosen just for the sake of responding to the question. Therefore, the length of the time for a testing session and diversifying user types should be carefully planned in future study.

7.1.2 Main reasons of the importance of the criteria to CSCL UE

The findings in section 6.4.1 provide some valuable insights about the reasons of the importance of the criteria in each dimension to CSCL UE. Overall, TG gave more reasons about the importance of the criteria to UE than FU did. More participants in TG responded to the questions (see Table 6-22 on pg 110 in Section 6.4.1.7). This could be because the participants in TG were sitting beside the facilitator and completed the questionnaires during the testing sessions whereas the participants in FU filled in the questionnaire during their spare time and some of them might not take this task seriously. Therefore, on the whole the quality of the responses from TG could be better and more complete than that of those responses from FU, although many responses from FU were valuable.

The following sections will discuss the findings about each dimension.

7.1.2.1 Effectiveness Dimension

This dimension has four criteria: Completeness, Visibility, Organisation/Design, and Navigability. The participants in both groups believed that the four criteria in this dimension would determine if a system is usable, accessible, comfortable, and attractive to the user and whether the system is user friendly and easy to use. And they also would affect the time for completing a task, trust building, and user experience (see Section 6.4.1). Both groups agreed this dimension would affect the time of completion of a task.

However, there was a slight difference between the two groups in assessing the relative importance, i.e. TG were more concerned about a system UI and understandability, whereas FU cared more about the ease of use and simplicity of a system. This might be because the participants in TG had to finish the testing task around 20 minutes and they only had a short time to experience the interface and layout design of OJS whereas the participants in FU had much longer time (more than a month) for using OJS to complete their group assignments. The participants in FU had gained more experience in using OJS than TG had.

7.1.2.2 Efficiency dimension

This dimension has four criteria, i.e. Speed, Effort, Familiarity/Consistency/Standards, and Flexibility/Adaptability/Configurability. The findings in Section 6.4.1.2 prove that both groups believed the criteria in the efficiency dimension are important because the features as defined by the criteria would impact ease of use, helpfulness, and reliability of a system. By having these features as part of a given system, users would quickly become familiar with the system's UI, and would be able to use it, complete a task, and get outcomes quickly with minimal (if any) outside support.

7.1.2.3 Collaborativity dimension

The Collaborativity dimension has eight criteria: User Management, Awareness, Communication, User Control/Moderator & Teacher control, File/Content Sharing & Management, Process Tracking/Automated Notification, File/Content Protection, and Security. They reflect the requirements of a collaborative learning and teaching system as suggested by Wolz et al. (1997) and Graham and Misanchuk (2004) (see Section 2.2.2). Some statements about synchronous communication were removed during the trial testing due to they were inapplicable to OJS as explained in Section 4.3.2.

The thoughts expressed by the participants in both groups and presented in Section 6.4.1.3 illustrate that the eight criteria are essential features of a CSCL system for group activities because they impact upon team performance and speed, group cohesion, and collaborative communication, and ultimately affect whether users can use a system to successfully complete team tasks and learn collaboratively, and are happy to use it.

7.1.2.4 Error tolerance

This dimension has two criteria, i.e. Error Rate and Error Prevention. The findings in Section 6.4.1.4 show that the participants believed that it is inevitable that people make errors. The two criteria for error tolerance would ensure that users can foresee/predict errors, decide what they should do when errors occur, and then consider how to avoid making similar mistakes to save time and resources. Moreover, they would affect the willingness of users to use a system and how easy a system is to use.

7.1.2.5 Universal accessibility

Universal accessibility dimension initially had three criteria and were removed after trial testing because they were not applicable to OJS and the usability testing in this case study as discussed in Section 4.3.2. So, this dimension has only one criterion, i.e. "support different users with different levels of IT expertise". Different people in a collaborative environment have varying levels of IT skills. The findings suggest that a system supporting teamwork should ideally help a range of people use the system easily regardless of their prior IT skills; this criterion would therefore directly affect whether the system would be easy to learn, its subsequent popularity and its cost efficiency (more details shown in Section 6.4.1.5).

7.1.2.6 Satisfaction

This dimension has five criteria: Usefulness/Functionality, Learnability/Predictability/ Recognition/Memorability, Simplicity, Help/Documentation, and Aesthetic Design. The findings in Section 6.4.1.6 demonstrate that these five criteria define the core features and requirements of a system. If the criteria are met, a system will be simple and useful, easy to use and learn, helpful to save time when working on team tasks or group collaborative learning. The criteria will also impact on whether users want to use a system and what they think of the quality of the system's usability.

7.1.3 Summary

In general, all participants agreed that all criteria in the framework were either important or very important to a system usability and UE. The most common reasons of rating the criteria as important criteria can be summed up as follows: the criteria would impact on

users c*ompleting a team task* and the *Speed/Time spending* on completing teamwork and team learning, and if a system is *easy to use & learn* and has *core/key/essential features*, and further affect *User experience* (see Table 6-22 on pg110). However, the insufficient time allocated to a testing session and similar user types of the participants might have limited the range of responses made by them.

7.2 Research question 2

Is the proposed framework capable of evaluating CSCL system usability in practice?

The objective of this question is to find out if the outcomes of using the framework to evaluate a CSCL system (i.e. OJS in this case study) could identify the advantages and the disadvantages of the system usability in an educational environment.

The findings collected from Part 1 and Qc in Part 3 of Post-test questionnaire and some questions for interview show with what statements on the features of OJS usability the participants agreed or disagreed and what features of OJS usability they liked or disliked. The implications of the findings are interpreted below:

1) if they agreed or liked some features, then the features could be seen as the advantages of OJS usability;
2) If they disagreed or disliked some features, then the features could be seen as reducing OJS usability and the disadvantages of OJS usability, and could be even defined as problems or issues of OJS usability.
3) If the answers were undecided, then possibly the features to which the statements refer would need further testing or the features were between good and bad, in other words, they were neither advantages nor disadvantages but could be fair or acceptable features of OJS.

7.2.1 Advantages and disadvantages of OJS usability

The findings from the interview Q9 in Section 6.6.1.5 and Q1 in Section 6.6.1.1 and Q2 in Section 6.6.1.2 illustrate that 91.67% of TG (11 participants) were satisfied with OJS usability overall. They agreed that it was *"good for the purpose"* and *"easy to use",* and liked the email tools. The main advantages of OJS as identified by TG were some useful tools/templates for email, storing and sharing files and ideas, easy and simple layout, and role administration. But 58.33% of TG (9 participants) did not like the navigation and 25% or TG (3 participants) disliked the features of error prevention.

With reference to collaborative teamwork in OJS (Q2 in Section 6.6.1.2), eight participants (66.67%) in TG thought that it was a ‘good’ or ‘ok’ system for asynchronous collaborative teamwork that did not need real time communication, such as file editing, paper review/peer review, and so on. However, some participants did not like it because they believed that a system for collaborative teamwork would need more features and synchronous communication for team work and CL.

The findings from Part 1 of post-test questionnaire in Section 6.2 show with which statements participants agreed or disagreed. More statements were agreed by TG than by FU. Table 6-11 in Section 6.2.7 (on pg 96) shows that 23 statements were agreed or strongly agreed and 5 statements were undecided by more than 50% of TG while 17 statements were agreed or strongly agreed and none of the statements were undecided by more than 50% of FU. And 50% of the participants in each group strongly agreed or agreed to 37.21% of all 43 statements (16 statements) in the post-test questionnaire. These findings indicate that the majority of participants in both groups liked the features defined by the 16 statements. Those features that the statements refer to could be considered as the advantages of OJS usability.

Table 6-12 in Section 6.2.7 (on pg 97) lists 20 statements ranked as N/A by one or more participants in a group. The features reflected by the statements ranked as N/A could be unavailable in OJS and needed to be developed in future. The unavailable features defined by the four statements ranked as N/A by over 50% of TG are related to Help/Documentation (Q6.4.2- easy to access help documents and Q6.4.3-easy to switch between help function and work), Security (Q3.8.1 -secure storing teams' work files), and File/Content Protection (Q3.7.1 - warning message for file sharing), while the unavailable features defined in the two statements ranked as N/A by over 50% of FU are related to User Management (Q3.1.1 -Manage user account and Q3.1.2 -Editor able to assign jobs). However the findings were incorrect because in fact, an editor was able to enrol users and assign roles to users and the help/documentation was available on the OJS website. That is, these features referred by Q3.1.1 and Q3.1.2 and Q6.4.2 were available in OJS. The reason for the participants in FU to give these responses might be that the role assignment was done by the lecturer of teaching CC. They were assigned as reviewers, authors and section editors in OJS and did not work on user/role assignment, but were given the OJS’s help manual when they were working on their group assignments. In contrast, the participants in TG worked as the Editor, and were required

to assign reviewers but they did not have time to browse help/documentation provided in OJS site during their testing sessions.

Table 7-1 next page summarises the advantages and disadvantages of OJS drawn from the findings from both post-test questionnaire and the interviews. The features that are reflected by the 16 statements and were selected by majority of both groups are highlighted in turquoise colour. For example, the advantages of OJS in Collaborativity dimension shown in the table are: User Control/Moderator & Teacher control (Q3.4.2- User control), File/Content Sharing & Management (Q3.5.1- easy file upload, Q3.5.2- easy file retrieval), Process Tracking/Automated Notification (Q3.6.1 - notification of task and Q3.6.2- show task status), security (Q3.8.2 - logon required). Three disadvantages defined by the participants are error prevention, unclear information and poor user interface (Visibility), and navigation (Navigability).

The responses to the open-ended question (Qc) asking for *"comments about each dimension of the system (OJS)"* in Part 3 of Post-test questionnaire provided additional thoughts about each dimension. Next the advantages and the disadvantages of OJS CL usability will be discussed in an order of the six dimensions. (The findings are shown in Section 6.4.2).

Table 7-1 The advantages / disadvantages defined by more than 50% of a group

Dimensions	Criteria (features)	Not available		Disadvantages		Fair / acceptable		Advan-tages	
	Groups	**TG**	**FU**	**TG**	**FU**	**TG**	**FU**	**TG**	**FU**
Effectiveness	Completeness (1.1.1)							✓	✓
	Visibility (1.2.1)			✓				✓	
	Organisation/Design (1.3.1 - User Friendly/Familiar)					✓			
	Organisation/Design (1.3.2 - Logical Steps)								✓
	Navigability (1.4.2 - clearly points me to the next step/task in a workflow)			✓					
Efficiency	Speed (2.1.1 - Work on tasks efficiently)							✓	✓
	Speed (2.1.2 - Speed of system)							✓	✓
	Familiarity/Consistency/Standards (2.2.1 - Familiar user interface)					✓			
	Familiarity/Consistency/Standards (2.2.2 - Consistent Layout)							✓	✓
	Effort (2.4.1 - no need to remember information)							✓	
	Effort (2.4.2 - no need for prior knowledge)							✓	
	Effort (2.4.3 - no need for technical support)					✓		✓	✓
Collaborativity	User Management (3.1.1 - Manage user account)		✓					✓	
	User Management (3.1.2 - Editor able to assign jobs)		✓					✓	
	Communication (3.3.1)					✓			
	User Control/Moderator & Teacher Control (3.4.2 - User control)							✓	✓
	File content sharing & management (3.5.1 - easy file upload)							✓	✓
	File content sharing & management (3.5.2 - easy file retrieval)							✓	✓
	Process Tracking/Automated Notification (3.6.1 - notification of task)							✓	✓
	Process Tracking/Automated Notification (3.6.2 - show task status)							✓	✓
	File/Content Protection (3.7.1 - warning message for file sharing)	✓							
	Security (3.8.1 - secure for storing teams' work/files)	✓							
	Security (3.8.2 - logon required)							✓	✓
Satisfaction	Usefulness/Functionality (6.1.1 - system is fully functional)							✓	✓
	Usefulness/Functionality (6.1.2 - useful for teamwork)							✓	✓
	Usefulness/Functionality (6.1.3 - integrated functions)							✓	
	Learnability/Predictability/Recognition/Memorability (6.2.1 - easy to learn)							✓	✓
	Learnability/Predictability/Recognition/Memorability (6.2.2 - task easy to perform)							✓	
	Simplicity (6.3.1 - system is simple to use)							✓	✓
	Help/Documentation (6.4.2 - easy to access help documents)	✓							
	Help/Documentation (6.4.3 - easy to switch between help function and work)	✓							
	Overall (6.6.1 - Reliable system)					✓			
	Overall (6.6.2 - Satisfaction with the system)							✓	✓
	Error prevention			✓					

Note: The advantages / disadvantages defined by more than 50% of each group are highlight in turquoise colour.

7.2.1.1 Effectiveness dimension

The participants in TG agreed that OJS was simple and efficient, ease of use in terms of navigation and completing the teamwork of a peer review; TG also found that email was useful for group communication while the participants in FU believed that OJS was easy, and simple.

Navigation and visibility were identified as the two main weaknesses of OJS usability. TG's participants believed that improvement on visibility, readability, and navigability, such as hyperlinks/icons/menus, text-based web pages, layout structure, pop up screen, etc were required. For example, menus should be more visible; contents on web pages should be more readable and colourful. FU's participants made similar suggestions and would like to have better navigation, layout, menus, and content flow on web pages as those on web pages were confusing and not user friendly.

7.2.1.2 Efficiency Dimension

The participants in TG thought that OJS was easy and helpful, although links/icons and the structure of the system needed improvement. Similarly, the participants in FU gave some positive comments: OJS was good, easy, and simple, and efficient for users to complete a task; OJS provided email templates so the participants did not need to make a lot of effort to complete the task and could save time. But they needed help/assistance when they were not familiar with the system. Sitemap/headings should be provided on the top of web pages. So Effort, Familiarity/Consistency/Standards, and Flexibility/Adaptability/Configurability of OJS were judged not good enough and need to be improved.

7.2.1.3 Collaborativity Dimension

The participants in TG identified some advantageous features such as displaying submitted articles and reviewers' recommendations, showing the review progress status (e.g. articles in review or assigned or unassigned), as well as visibility of users' roles. Meanwhile, some negative features were noted such as confusing workflow and information, easily getting lost, weaknesses in security and awareness. The participants in FU commented that users can easily complete a collaborative task, e.g. communicating with other teammates via email, tracking a process, and managing files but OJS had only one communication channel – email, and user's files are not traceable. Generally speaking, the participants from both groups liked the following criteria to some degree: User Control/Moderator & Teacher control, File/Content Sharing & Management, email tools, and Process Tracking/Automated Notification. However,

they believed User Management, File/Content Protection, and Security were not applicable; Awareness and Communication needed to be further enhanced; and OJS should have more features/capabilities supporting synchronous communication.

7.2.1.4 Error Tolerance Dimension

Some participants in TG commented that OJS provided warnings or error messages by confirmation dialogs to advise user how to withdraw or cancel incorrect or irreversible actions when some mistakes were made. But many participants thought that OJS did not provide proper warnings or error alerts when they made mistakes. Some participants' opinions from FU were similar as the participants' from TG. FU's participants stated that OJS did provide some information about what should be done to avoid some mistakes but it did not have the reverse actions capability. So, these features would be the downsides of OJS usability. The participants in both groups suggested that OJS should provide users with error alerts or additional functions like undo, reverse, or instructions on how to perform a required action in order to avoid user making mistakes.

7.2.1.5 Universal Accessibility Dimension

Some participants in TG did not like the feature provided by OJS and gave negative comments: some navigational elements like links and buttons did not have a universal design and were inconsistent. But a few participants in FU thought it was easy and that the feature was fine. However both groups believed that this dimension of OJS usability needed to be improved.

7.2.1.6 Satisfaction Dimension

The majority of participants were satisfied with using the system and more participants in FU were satisfied with the system and gave more positive comments on OJS than the participants in TG did as shown on Table 6-24 (on pg 116). The reason for the higher satisfaction rate from FU could be that they worked on their assignments during 2 months and had more time to explore and understand the features of OJS whereas each participant in TG was given around 20 minutes for working on the testing task as mentioned before.

The participants in TG were satisfied with using OJS because they could use it to fulfil the testing task easily and simply with little help. But they did not think that OJS provided them with group email, a simple system structure, some good tool-tips, integrated help, or a strong search engine. Moreover, page layout and design, and text-based pages were not user friendly. About Help/Documentation, 75% of TG (9 persons) (shown on Table 6-12 on pg 97) considered it as not applicable.

Similarly, FU agreed that OJS was easy and simple to understand, learn, and use for collaborative tasks, especially for submitting and reviewing articles. It is useful to have an online review function. Email tools, such as email templates and a hyperlink to a file to be reviewed on an email could save users' time when completing a team task. But the web pages could not be configured based on users' needs. Some participants in FU believed that some improvement on flexibility, colour design and documentation should be undertaken in future.

Overall, the features of OJS CL usability defined by Usefulness, Learnability/ Predictability/ Recognition/ Memorability, and Simplicity were considered to be pretty good, while the features reflected by Flexibility, Help/Documentation and Aesthetic Design were either not available or not designed and developed properly.

7.2.2 Summary

OJS has more advantages than disadvantages. The majority of participants were satisfied with using the system when they worked on the testing task or on their assignments. Both TG and FU groups had identified 16 advantages and 3 major disadvantages. Some features reflected by the following criteria need to be improved in future are: visibility, readability, and navigability in the Effectiveness dimension; Effort, Familiarity/Consistency/Standards, and Flexibility/Adaptability/Configurability in the Efficiency dimension; User Management, File/Content Protection, Awareness, Communication, and Security in the Collaborativity dimension; error prevention in the Error tolerance dimension; universal interface design in the Universal accessibility dimension; Flexibility, Help/Documentation and Aesthetic Design in the Satisfaction dimension.

7.3 Research question 3

What should be improved in terms of the future study in developing a framework for CSCL system UE in educational settings?

This question intends to look into the issues and problems occurred during the case study and then find out what areas of the study should be improved in future study.

Several issues and problems occurred during the study. They can be categorised into three types: people-related, system-related, and study-related. This section discusses the

problems happened during the case study and then highlights the main areas that should be improved in future study in developing a framework for CSCL system UE.

7.3.1 People-related issues and problems

People make mistake inevitably. During the case study, several issues and problems occurred. The main issues or problems are unfamiliarity with OJS, problems occurred to the CC class, and the participants recruited from one class.

7.3.1.1 Unfamiliarity with OJS

Only one 30-minute demo was presented to an evening class before conducting the testing. The findings in Section 6.5.1.5 illustrate that most participants in TG had never used OJS before taking part in the testing. They also had little knowledge about peer review process, so they had to follow the task script when working on the task. However, some participants did not want to read the task script but instead just worked on the testing task based on their intuition. The end result was that these participants got 'lost', made mistakes, and had to spend some time on figuring out how to complete a given step/subtask. This also happened to the students in FU group when doing their group assignments. The participants in FU were given a help manual by email but some of them did not read it carefully and ended up with having problems when completing one round of peer review. So, the participants' experience in the testing and using OJS might have affected their responses to the interviews and questionnaires.

The researcher was not familiar with OJS and was not able to customise the OJS journal site for usability testing and to make it look more colourful and have a better layout. The original page design, layouts, themes, and tools built in OJS software were adopted without any modification and applied to the journal site for testing. The web pages were full of text. This might be why some participants thought that the testing journal site was not set up completely.

7.3.1.2 Problems occurred to the CC class

Several problems happened when the students in the CC class used the OJS journal site to work on their group assignments (more details shown in Appendix 27). For example, some students uploaded their assignments to a wrong location, some failed to submit the right files, and others made mistakes during the peer review process. Consequently, some students struggled with completing their peer reviewing on time and disliked the system very much. This could have affected their responses to the Post-test questionnaire. In general, these problems were mainly related to students having little

knowledge in the peer review process and OJS system. And they did not read the given manual and instructions carefully.

7.3.1.3 Participants recruited from one class

"*Subjective satisfaction ratings are not a very telling usability measure because users tend to give generous scores even when they have great difficulty using a design. Another reason is that users often don't know how poorly they performed when they tested a site*" (Nielsen & Loranger, 2006). Moreover, the researcher was unable to recruit various user types and then to gather a wide range of responses/views on the framework and OJS UE due to the time limitation of the part time study. The participants were mainly recruited from the students in the CC class in SCMS at AUT. Majority's expertise levels in using the CSCL system and OJS were either level 1 "Novice" or level 2 "Beginner" according to the findings in Section 6.1.2. As a result, their responses might not reflect their true real usability measure and could be similar. Some responses were related to the system web page interface rather than delving deeper into the collaborative learning and teaching aspects. Some like "yes" or "agree" might have been a guess or chosen just for the sake of responding to the question as discussed in Section 7.1.1. Although the participants' responses had provided some valuable insights on OJS usability and the importance of the criteria in the framework, some of them were not useful for a deep analysis and drawing a concrete conclusion.

In summary, people-related issues and problems happened during the study. The ways of improvement suggested from this study are providing participants with training sessions, and help information or manuals prior to conducting a testing, recruiting participants from both students and staff with different expertise levels, and getting a large sample size.

7.3.2 System-related issue

There is an issue related to the limitation of OJS which was used as a CSCL system.
OJS is a free system for online journal publishing. It does not have some of the tools and utilities normally used in synchronous teamwork, e.g. instant messaging, chatting, and pop up alerts when a teammate logs on. Email is the main communication channel in OJS. The findings in Section 6.6.1.3 show that several participants believed that OJS fit in with its purpose – journal publishing system but not with the collaborative learning and teaching. The collaborative teamwork on OJS had to be asynchronous. OJS was not a complete CSCL system. Consequently, the research was limited to the peer

review process that could be seen as a kind of asynchronous learning and study activity. Therefore, the criteria for measuring synchronous teamwork and real time communication were not included in the framework and tested in this study because the features represented by them were not available on the OJS testing site. This also had affected the participants' responses to some questions about the Collaborativity and Satisfaction dimensions. Many responses to some questions about these two dimensions were "Not Applicable" or "No responses".

In order to overcome the limitations explained above, it is suggested that a selected system should fit well into the context – collaborative teaching and learning in educational settings; and should be available most time in the course of a research.

7.3.3 Study-related issues

Some issues are related to the case study itself: only one system was tested due to time constraint, the quality of the research were negatively affected by a small sample size, there was a step missing in the testing design, and it was difficult to design a collaborative testing task and set up a real collaborative working environment.

7.3.3.1 Time constraint

The study had been a 2 year part-time study. The participants were recruited from the postgraduate students who studied the CC course and used OJS as a CSCL system to work on their group assignments. OJS was the only system used for user testing in this study and could be accessed by the potential participants (the students and staff in SMCS) only when CC course was taught. It was impossible to use the framework to evaluate OJS more than once as the CC paper was taught in one semester during the period of the study.

7.3.3.2 The sample size and internal reliability

The results of reviewing literature in Section 3.4.1.1 on pgs 36 -37 show that a general rule for the optimal sample size for user testing can be '4±1' (Rubin & Chisnell, 2008) or '10±2' participants (Hwang & Salvendy, 2010). "*Think Aloud requires nine test users to reach 80% overall discovery rate*" (Hwang & Salvendy, 2010). However identifying the usability deficiencies of a CSCL system was not the objective of this study. So, applying these rules to this study was inappropriate.

Regarding data internal reliability analysis, as explained in Section 5.5, *"a research has at least 10-15 participants per variable"* (Field, 2005). The number of the criteria in the framework is 25 so the size of the sample should be at least 250. However, the sample size in this study was very small: 18 participants were recruited, 12 testing sessions were conducted, 12 TS, 12 ODCS and 25 each of the two questionnaires were completed. In addition, some participants, particularly those from FU left many questions with no answers. Those participants' data with no responses were treated as missing data and excluded from the analysis in SPSS. This further reduced the size of the data sample and affected data reliability and the research quality.

7.3.3.3 A step missing in the testing design

"*One drawback of user testing is the difficulty in training users within a limited amount of time to master advanced features of a Web application. This can lead to shallow conclusions, in most cases only related to the simple application features*" (Matera, Rizzo, & Carughi, 2006). Most participants in TG had never accessed OJS until they came to the testing sessions. Within an hour, they had to complete three subtasks – the testing task, questionnaires, and an interview. They might not have enough time to explore the system thoroughly and to think about the system usability.

In two testing sessions, the researcher deliberately gave the participants a few minutes to read the task script and then let them start working on the task. The outcomes of the two testing sessions were better than that in the previous sessions. They spent less time on completing the task and made fewer mistakes than others who did not read the script before carrying out the task. So, it would be better if the task script was sent out to the participants and the participants were given access to the OJS site before they came to the testing sessions. This was also suggested by a few participants in the interviews (see Section 6.6.3). But this step was not included in the testing design and the testing plan.

7.3.3.4 The difficulties of designing a collaborative testing task and setting up a collaborative working environment

"*Failure to reproduce a real environment where the application is to be used may lead to unrealistic results and conclusions*" (Matera, et al., 2006). Peer review in OJS is an asynchronous process. At the beginning of designing the testing plan and the testing task, the researcher felt very hard to simulate a collaborative working environment where participants could work on different computers at different time while they could be observed by one facilitator. Another difficulty was how to design a collaborative team task which could be completed in OJS in a short time. A commonly acceptable

time for a testing session is an hour. If the time is longer than an hour, potential participants would lose interest in participating. If a testing session is split into different sub-sessions at different time, then time arrangement with participants would be very difficult and even impossible.

After several discussions with the supervisors and the two trial testing sessions, the testing plan was eventually set up, i.e. a testing session included a participant completing a testing task, a task sheet, two questionnaires, and an interview within an hour (see Section 4.2.2 and 4.3.3). The collaborative team task was peer review on OJS, was simplified and had to be completed around 20 minutes by using one computer in one room which was not an actual collaborative working environment. During a testing session, no asynchronous discussion and communication actually happened. This could have affected what the participants in TG thought of the collaborative feature in OJS and their understanding OJS usability, and further impacted their responses to the questionnaires, particularly to the questions about the Collaborativity dimension. In future study, the established testing environment should be close to the real working environment and the testing task should be designed as a real collaborative learning and teaching task.

7.3.4 Summary

Several problems and issues occurred during the case study and can be categorised to three types: *people-related issues* including participants unfamiliar with OJS and the peer review process, errors occurred to the CC class, and participants recruited from one class only; *system-related issue*, i.e. OJS was not a complete CSCL system so some criteria for evaluating some features of real time communication and collaborative teamwork were not included in the framework and tested; and *study-related issues and difficulties*, including time constraint, small sample size, a step missing in the testing design, and the difficulties in designing a collaborative testing task and setting up a collaborative working environment. Those problems and issues had negatively impacted on testing the proposed framework in evaluating OJS usability, participants' responses to the questionnaires and the interview, and further impacted the quality of the research.

7.4 The framework for CSCL system UE

7.4.1 Framework and its criteria

As discussed on the research question 1 in section 7.1 and the question 2 in section 7.2, the findings drawn from Part 2 in Section 6.3, Part 3 - Qb in Section 6.4.1, and as well as the interviews in Section 6.6.2 demonstrate that the criteria in the framework were important to a system usability and UE, and should be included in the framework. The findings from Part 1 in Section 6.2 and Part 3 – Qc in Section 6.4.2 and the interviews in Section 6.6.1 prove that the framework was capable of discovering the advantages and disadvantages and as well problems of OJS CL usability. The findings collected from the interview question 5 ("*Do you think the framework covers mostly important criteria? Why do you say so? What else need to be included in the framework?*") in section 6.6.2.2 illustrate that the framework consisted of a wide range of important criteria. One suggestion is to add a new criterion about backup to the framework while another suggestion is to reduce the number of the criteria. So, the criteria are important and should be kept in the framework for CSCL system UE.

Next the findings of IR and IRR analysis for Part 2 of post-test questionnaire will be reviewed to further check if modification of the framework is needs.

7.4.2 Implication of IR analysis and IRR analysis – ranking importance

The findings from the IR analysis in Section 5.2.2 show that the internal consistency of both groups' rankings of the importance of the criteria for each of the three dimensions (Satisfaction, Collaborativity, and Error tolerance) was either questionable or acceptable, whereas, the internal consistency of both groups' rankings of the importance of the criteria for each of the other two dimensions (Effectiveness and Efficiency) was unacceptable. Therefore, the internal consistency of the importance levels of the criteria in the Effectiveness and Efficiency dimensions might need further investigation by getting larger responses or need to be restructured in future. However, none of the criteria that should be removed from these two dimensions were identified.

The findings from IRR analysis in Section 5.3.2 illustrate that TG's results were different from FU's (see Table 5-8 on pg 83). As discussed in Section 5.4, this could imply that the participants did not have the similar views on the importance levels of the criteria in the four dimensions. For ranking importance levels of the criteria in the Collaborativity dimension, within a group, TG's responses were fairly consistent but

FU's responses were inconsistent. For Satisfaction dimension, the responses in TG were excellently consistent but the responses in FU were inconsistent. For the Effectiveness and Efficiency dimensions, the participants' responses were poorly consistent within TG and fairly consistent within FU.

About the Effectiveness and Efficiency dimensions, the consistencies of IR analysis and IRR analysis were poor or fair. This could further imply that IR inconsistency of the importance levels of the criteria in these two dimensions might be unreliable. So, the first priority of future study should not be restructuring the two dimensions but further investigation in the importance of the criteria in these two dimensions by getting larger responses i.e. by getting a larger data sample.

As discussed in Section 5.4, three criteria that were identified to be removed from their dimensions are "Communication" (Q3a3) and "Security" (Q3a8) in the Collaborativity dimension and "Help/Documentation" (Q6a4) in the Satisfaction dimension. But the findings shown in Table 6-19 (on pg 103) indicate that they were very important or important to CSCL system usability and these three items should be included in the framework.

As stated in Section 5.5, the outcomes of the IR and IRR analysis could be seen as indicative and inconclusive. Normally the ratio between the sample size and variables should be *10-15 participants per variable* (Field, 2005) but the data sample size is very small - only 25 participants in this study. The size of the data sample could be too small to produce cogent evidences for drawing a conclusion. Therefore, the researcher believe that at this stage, the criteria should be included in the framework and should be retested in future study which should have a large number of participants who apply the framework to evaluate a CSCL system.

7.4.3 Implication of IR analysis and IRR analysis – ranking agreements

The outcomes of IR analysis shown in Section 5.4 illustrated that the internal consistencies in different dimensions were various. Regarding the agreements on the statements of the criteria used for evaluating OJS CL usability, the internal reliabilities in three dimensions (Effectiveness, Efficiency, and Error Tolerance) were acceptable and good (agreed by both groups) and the internal reliability in the Satisfaction dimension is excellent (agreed by FU only). However, the agreements on the statements of the criteria in the Collaborativity dimension were internally inconsistent so the

statements of the criteria might not completely reflect some OJS features in the Collaborativity dimension. The statements for this dimension may need to be retested by a larger number of participants or to be revised or restructured. In addition, Table 5-7 on pg 82 shows that five statements (three in the Collaborativity dimension and 2 in the Satisfaction dimension) might need to be retested, or revised, or removed from their dimensions in order to increase their dimensions' IR consistency.

According to IRR analysis in Section 5.3.1, overall the agreements among the participants within a group were consistent or acceptably consistent for the three dimensions (Effectiveness, Efficiency, and Collaborativity) but not consistent for the Satisfaction dimension. The rest two dimensions (Error tolerance and Universal Accessibility) were not included in the IRR analysis because the number of cases was too small.

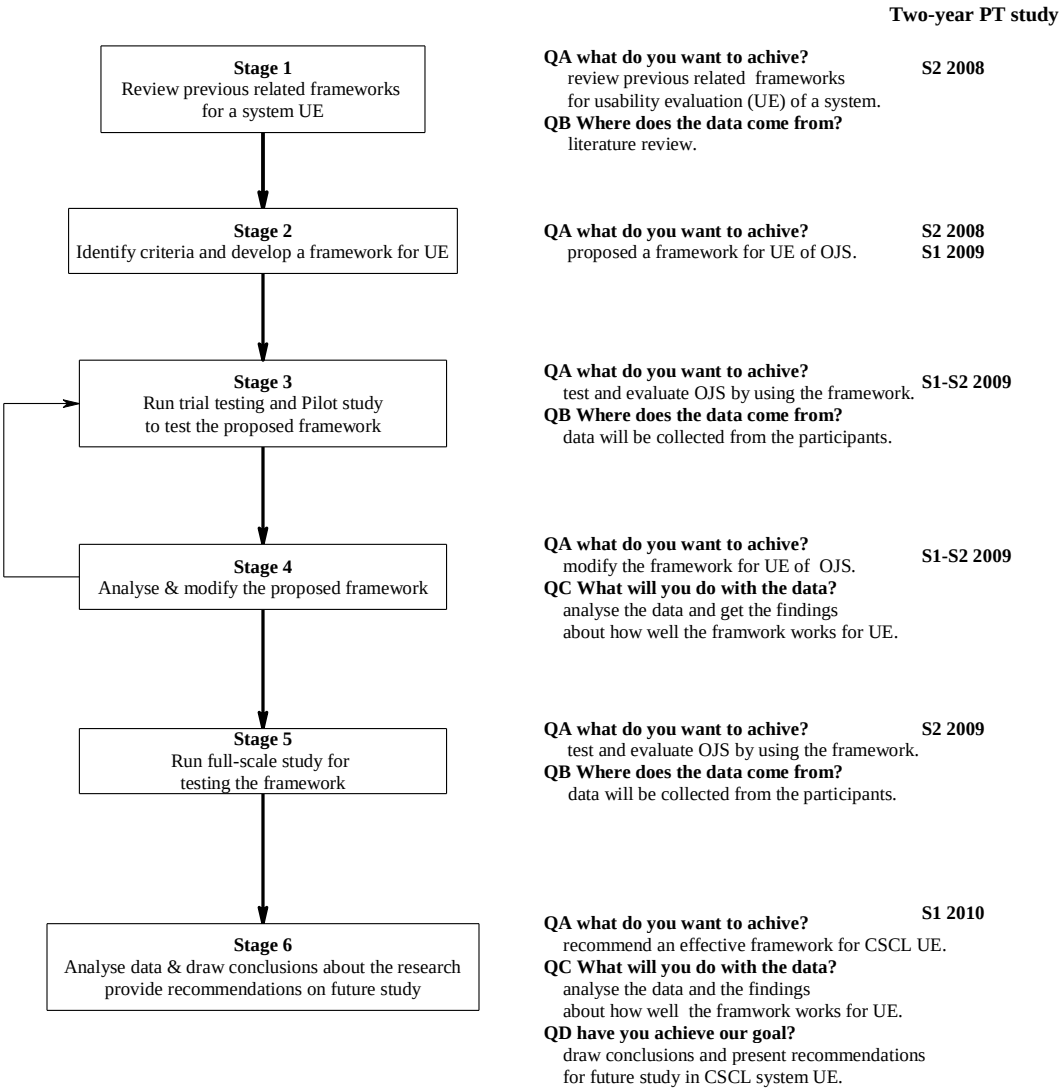
As explained in Section 5.5, the data sample in this study is too small. Moreover, for the Collaborativity dimension's IR analysis, only 3 cases' data from each group was processed in SPSS (see Table 5-1 on pg 75). Therefore, the outcomes of the reliability analysis based on the data about ranking agreements on the criteria in the five dimensions (Universal accessibility excluded) should be referential and inconclusive. Further study in testing the proposed framework in CSCL system UE with a large number of responses would need to be considered.

7.4.4 Summary

The findings from interviews and post-test questionnaire prove that all criteria in the framework were important to system usability and usability evaluation, and the framework is capable of identifying advantages and disadvantage of OJS CL usability. Because the size of the data sample was too small to produce cogent evidences for drawing a concrete conclusion, the findings of the IR and IRR analysis do not prove what criteria or statements should be removed from the framework but suggest that further testing the framework with a large sample size would need to be considered.

7.5 Application of the Methodology in this study

Figure 7-1 The application of the methodology in the study



The research employed the methodology designed in Section 3.3.1 and smoothly completed the five stages before coming to the last stage - stage 6. Figure 7-1 above shows the actual progress of this study. It took four and a half months to complete one round from stage 3 "trial testing" to stage 4 " evaluating and modifying the framework", then back to stage 3 "Pilot study" and stage 4, and then moved to stage 5 "full-scale study". It did not go repeatedly through stage 2 to 4, and back to 2 or 3 and then to stage 4 – that was defined in Section 3.3.1, i.e. "*stages 2, 3, and 4 can be part of an iterative process where any modifications to the chosen framework will be incorporated*". The reason was that the testing only could be conducted when the CC paper was taught (in S2 2009) and the study timeframe was 2 years. In summary, this methodology had guided the research to achieve its goals.

7.6 Summary

This chapter has discussed the findings presented in chapter 6 to answer the three research questions. In relation to question 1: most participants gave similar answers, that is, the criteria were important to CSCL system usability and UE, and the framework consisted of a wide range of criteria for UE. The findings related to question 2: some advantages and disadvantages of OJS were identified according to the findings found from testing, questionnaires, and interviews. The outcomes of the findings from the three resources are quite similar. Overall Effectiveness, Efficiency and Satisfaction were given more positive rankings and comments while Error tolerance, Universal accessibility, and some features of Collaborativity were given more negative rankings and comments. Majority of participants were satisfied with using OJS in the testing sessions and for completing group assignments and peer reviews. Some issues and problems of the research were identified in relation to question 3. The most serious concerns are the small sample size for this research, the limitation of OJS for CL activities in an educational context, and the difficulty of designing and conducting a collaborative task for usability testing, as well as the limited time issue and people-related issues.

About the proposed framework for CSCL system UE, the findings prove that the criteria are important and should be included in the framework. Because the sample size was very small, the findings of the IR and IRR analysis were just inconclusive. A further study with a large number of responses in retesting the framework i.e. using the framework to evaluate a CSCL system's usability would need to be considered.

The methodology provided a guideline for implementing the OJS usability testing. The completion of the study has proved it is a proper methodology for this study.

Chapter 8

8. Conclusion and recommendations

8.1 Conclusion

A review of related research has found that there was a gap between developing a framework for CSCL system UE in educational collaborative teaching and learning settings and developing a framework for groupware UE in general settings. This study had adopted three essential dimensions (Effectiveness, Efficiency, and Satisfaction) as defined by ISO 9241-11, and then added 3 additional dimensions - Collaborativity, Universal accessibility, and Error tolerance to form a framework for CSCL system UE.

The framework consisted of 25 criteria that were further grouped into six dimensions. It then was used as a guideline to develop a post-test questionnaire for a case study. The participants tested OJS and then were given the questionnaire to evaluate its usability. The findings from the post-test questionnaire and the interview proved that the framework had included the important criteria for system usability evaluation and was capable of evaluating an asynchronous collaborative learning system, i.e. OJS, and identifying the system's advantages and disadvantages as well as its associated problems and issues. The findings of the internal reliability analysis and the inter-rater reliability analysis did not get sound evidences for identifying what criteria and statements should be removed from the framework because the sample size was too small to produce cogent evidences for drawing a concrete conclusion.

The methodology for this research was designed as a two year case study in UE (completed by the researcher on a part-time basis) with six stages. The UE methods defined in this study consisted of user testing (Think Aloud Protocol and user observation), interview, and questionnaire. The completion of the case study has demonstrated that the methodology was easy to implement and appropriate for this type of UE study.

The limitations and issues from this study are that OJS used to test the framework was not a complete CL system so some criteria in the framework were not tested and some criteria for evaluating real-time collaborative teamwork were not included in the framework. Furthermore, the sample size was too small and there were difficulties in

setting up a collaborative working environment. Also designing a suitable collaborative teaching and learning task was quite difficult.

Overall this study has achieved the three objectives of the research as defined in the original proposal of the research, that is: it successfully identified important criteria and proposed a framework with six dimensions for evaluating CSCL system usability, designed a collaborative learning task to test OJS and then employed the framework to complete the OJS usability testing and evaluation, and next identified the advantages and disadvantages of OJS. Then it defined the issues, difficulties, and problems occurred during the case study and identified the areas that should be improved in further study in developing a framework for CSCL system UE. The study has contributed towards narrowing the gap between developing a framework for CSCL system UE in educational settings and developing a framework for groupware in general settings.

8.2 Recommendations

When doing a research on development of a framework for CSCL system UE in a Case study, one should consider the following:

1) **Sample size:** a large sample size will ensure that the research outcomes will be more accurate and reliable. One should recruit participants from both students and staff with different IT skills and get a large sample size based on the rule – "*10-15 participants per variable"* (Field, 2005);
2) **The system to be selected for usability testing**: one should select a CSCL system that has been widely used in an educational setting for collaborative learning and teaching. This will ensure that its main features and tools have already met the requirements of a CSCL system. The criteria in the framework can then be used as intended i.e. for assessing the CSCL system.

   The potential participants may have chances to access a used system and have got the skills in working on the system. If not, one should provide training sessions or help and documentation to the participants and give them the access to the system before the testing. This would help overcome the drawback (unfamiliarity) and also likely reduce the number of 'Not Applicable' responses in questionnaires when participants are asked to rate features that are not available.
3) **The testing environment and the testing task**: "*Failure to reproduce a real environment where the application is to be used may lead to unrealistic results*

and conclusions" (Matera, et al., 2006). The environment is where a participant completes a testing task in a CSCL system. It should be set up as the same as the real working environment so the criteria in the framework will be fully applied to the system UE. One challenge is to design a suitable yet inexpensive collaborative learning task that can be managed and monitored easily by a researcher and also can be completed by a participant in a short time span.

4) **A framework and the context of a system**: the goals and the usability features which a system has in one context may be different from those that it has in another context. One should design and develop a UE framework which is adaptable and flexible enough to fit into the required context.

8.3 Future study

> *"Collaborative learning environments, pose new challenges and questions concerning the effective evaluation process." "definition of suitable methods for usability evaluation of learning technology should take place after considering the effect of usability in system utility, i.e. the expected learning outcome."* (Tselios, et al., 2008)

This study aimed to develop a framework for evaluating CSCL system usability in educational settings. However, the proposed framework was only tested on one system (OJS) that was not a complete CSCL system. Therefore, the following aspects should be considered and further investigated in future study.

Diversifying user types and increasing the number of participants: the sampling population should be diversified. The future study should recruit a large number of participants with a range of IT skills (various user types) in the areas of using CSCL system and evaluating CSCL system usability if possible. This will provide an opportunity to collect a wide range of views on CSCL system UE and the framework being tested, and to increase the data validity and reliability, and the quality of a study.

Testing several CSCL systems: a proper CSCL system used for testing the framework and criteria should be a system that has been already in use for collaborative learning and teaching courses or other programmes and available all the time to majority of students and staff in a tertiary environment. So it can be easily accessed and learned by potential participants. If there is a better fit

between the framework and the system under evaluation then it will be more likely that the framework can be refined and improved. The proposed framework in this study should be tested in two or more CSCL systems so the findings can be compared and the framework can be modified and customised to become an effective framework for CSCL system UE.

Refining the framework and weighting the dimensions: *"High level of effectiveness and efficiency in task execution does not contribute necessarily to the learning outcome. Thus, usability of learning technology should be redefined and related with its expected learning utility, which needs to be clearly determined."* (Tselios, et al., 2008) The framework would be ideal if it could enable a researcher to evaluate various learning utilities and to adjust the weighting of each dimension or criterion according to its relative importance to the context and the learning outcomes. So, evaluating a CSCL system's usability could be done more effectively. The proposed framework in this study is not a mature framework, and need to be further tested and refined. The future study should apply the framework to evaluate different CSCL systems, then identify what criteria should be removed and what else criteria should be added in, and next determine the weighting of each dimension in the framework.

To sum up, *"further research is required in this area in order to transform effectively these theoretical characteristics into design specifications and solid usability evaluation methods"* (Tselios, et al., 2008). Identifying an effective framework for CSCL system UE can be achieved but it still has a long way to go.

References

Abascal, J., Arrue, M., & Vigo, M. (2007). Human Computer Interaction Research in Web Design and Evaluation. In P. Zaphiris & S. Kurniawan (Eds.), *Human Computer Interaction Research in Web Design and Evaluation* (pp. 185-207). London: Idea Group Publishing. Retrieved 16 Jul. 2008

Ardito, C., Costabile, M. F., Angeli, A. D., & Lanzilotti, R. (2006). *Systematic evaluation of e-learning systems: an experimental validation*. Paper presented at the meeting of the Proceedings of the 4th Nordic conference on Human-computer interaction: changing roles, Oslo, Norway. Retrieved from http://portal.acm.org.ezproxy.auckland.ac.nz/ft_gateway.cfm?id=1182496&type=pdf&coll=ACM&dl=ACM&CFID=7843053&CFTOKEN=83891791 doi:http://doi.acm.org/10.1145/1182475.1182496

Barnum, C. M. (2011). *Usability Testing Essentials* Boston: Morgan Kaufmann. Retrieved from http://www.sciencedirect.com/science/article/B6Y3K-511965B-F/2/41182b4d96a7cac6fed531ff96d9384f. doi:10.1016/B978-0-12-375092-1.00004-0

Bennett, S. (2004). Supporting Collaborative Project Teams Using Computer-Based Technologies. In T. S. Roberts (Ed.), *Online collaborative learning: theory and practice* (pp. 1-27). Hershey PA: Idea Group Publishing. Retrieved 5 Jan 2009

Bevan, N. (1995). Human-Computer Interaction Standards. In E. Anzai & Ogawa (eds) (Chair), Symposium conducted at the meeting of the 6th International Conference on Human Computer Interaction, Yokohama

Borgman, C. L., & Rasmussen, E. (2005). Usability of Digital Libraries in a Multicultural Environment. In Y.-L. Theng & S. Foo (Eds.), *Design and usability of digital libraries : case studies in the Asia-Pacific* (pp. 270-284). Hershey, PA: Idea Group Pub. Retrieved 14 Jul. 2008

Brinck, T. (2005). *Groupware - Introduction, Applicatons, Design Issues, and links*. Retrieved 15 Nov. 2008, from http://www.usabilityfirst.com/groupware/index.txl

Brooke, J. (1996). *A quick and dirty usability scale*. Retrieved 15 Oct., 2008, from http://www.dis.uniroma1.it/~bertini/ipc04/materiale/quest/sus.pdf

Bryman, A., & Cramer, D. (2009). *Quantitative Data Analysis with SPSS 14, 15, & 16. A guide for Social Scientists*. London and New York Routledge.

Case, M. M., & John, N. R. (2007). PUBLISHING JOURNALS@UIC. *ARL: A Bimonthly Report* (2 5 2 / 2 5 3), pp12-15. Retrieved from http://www.arl.org/bm~doc/arl-br-252-253-uic.pdf

Chamberlain, K. (1995). *What is grounded theory?* . Retrieved 14 May, 08 from http://kerlins.net/bobbi/research/qualresearch/bibliography/gt.html

Chattratichart, J., & Brodie, J. (2004). *Applying user testing data to UEM performance metrics*. Paper presented at the meeting of the CHI '04 extended abstracts on Human factors in computing systems, Vienna, Austria. Retrieved from http://portal.acm.org.ezproxy.auckland.ac.nz/ft_gateway.cfm?id=986003&type=pdf&coll=ACM&dl=ACM&CFID=7951349&CFTOKEN=87599161 doi:http://doi.acm.org/10.1145/985921.986003

Chýla, R. (2007). What open source webpublishing software has the scientific community for e-journals? Symposium conducted at the meeting of the CASLIN 2007 Stupava (Slovak Republic). Retrieved from http://eprints.rclis.org/archive/00010870/01/rchyla_open-source_e-journal_systems.pdf

Cohen, A., & Court, D. (2003). Ethnography and case study: a comparative analysis. *Academic Exchange Quarterly, 7*(3), 283-287.

Collis, J., & Hussey, R. (2009). *Business research : a practical guide for undergraduate & postgraduate students* (3rd ed.). Basingstoke: Palgrave Macmillan.

Cortina, J. M. (1993). What Is Coefficient Alpha?: An Examination of Theory and Applications. *Journal of Applied Psychology, 78*(1), 98-104. doi:10.1037/0021-9010.78.1.98

Cronbach's alpha. Retrieved 3 May, 2010, from http://en.wikipedia.org/wiki/Cronbach%27s_alpha

CUD. (2008). *Universal Design Principles*. Retrieved 10 Oct. 2008, from http://www.design.ncsu.edu/cud/about_ud/udprinciplesshtmlformat.html#top

Cyzyk, M. (2007). A Survey and Evaluation of Open-Source Electronic Publishing Systems (ppt file). Retrieved from https://wiki.library.jhu.edu/download/attachments/22964/cniFall2007.ppt?version=1

Cyzyk, M., & Choudhury, S. (2008). A Survey and Evaluation of Open-Source Electronic Publishing Systems (white paper) [Library Digital Programs]. Retrieved from https://wiki.library.jhu.edu/download/attachments/22964/Open+Source+ePublishing+Systems+White+Paper.pdf?version=1

D'Hertefelt, S. (1999). *Observation methods and tips for usability testing*. Retrieved 10 Apr., 2009, from http://www.interactionarchitect.com/knowledge/article19991212shd.htm

Damianos, L., Hirschman, L., Kozierok, R., Kurtz, J., Greenberg, A., Walls, K., et al. (1999). Evaluation for collaborative systems. *ACM Comput. Surv., 31*(2es). doi:http://doi.acm.org/10.1145/323216.323362

Damianos, L. E., Drury, J., Fanderclai, T., Hirschman, L., Kurtz, J., & Oshika, B. (2000). Evaluating Multi-party Multi-modal Systems Symposium conducted at the meeting of the LREC2000 Proceedings: Session SO6 - Recognition, Athens Greek. Retrieved from http://gandalf.aksis.uib.no/non/lrec2000/pdf/368.pdf

Davey, L. (1991). The application of case study evaluations. *Practical Assessment, Research & Evaluation, 2*(9). Retrieved from http://PAREonline.net/getvn.asp?v=2&n=9

Dawson, J. W. (2006). *A holistic usability framework for distributed simulation systems* (Ph.D.). University of Central Florida, United States -- Florida. Retrieved from http://ezproxy.auckland.ac.nz/login?url=http://proquest.umi.com/pqdweb?did=1203572691&Fmt=7&clientId=13395&RQT=309&VName=PQD3233646)

DeBoard, D. (N/A). *Heuristic Evaluation Questionnaire (for online help)*. Retrieved 14 Feb, 2009, from http://www.stcsig.org/usability/resources/toolkit/heuristics-for-help.pdf

Deursen, A. J. A. M. v., & Van Dijk, J. A. G. M. (2010). Measuring Internet Skills. *International Journal of Human-Computer Interaction, 26*(10), 891 - 916. doi:10.1080/10447318.2010.496338

Dimitracopoulou, A. (2005). *Designing collaborative learning systems: current trends \& future research agenda*. Paper presented at the meeting of the Proceedings of th 2005 conference on Computer support for collaborative learning: learning 2005: the next 10 years!, Taipei, Taiwan. Retrieved from http://portal.acm.org.ezproxy.auckland.ac.nz/citation.cfm?id=1149293.1149309&coll=ACM&dl=ACM&CFID=12025639&CFTOKEN=38359002#

Dix, A., Finlay, J., Abowd, G. D., & Beale, R. (2004). *Human-computer interaction* (3rd ed.). New York: Pearson/Prentice-Hall.

Downey, L. L. (2007). Group Usability Testing: Evolution in Usability Techniques. *Journal of Usability Studies, 2*(3), 133-144.

Dumas, J. S., & Fox, J. E. (2008). Usability Testing: Current Practice and Future Directions. . In A. Sears & J. A. Jacko (Eds.), *The Human-Computer Interaction Handbook. Fundamentals, Evolving Technologies, and Emerging Applications* (2nd ed., pp. 1129-1149). New York: Lawrence Erlbaum Associates, Taylor & Francis Group. Retrieved 23 Aug 09

Dumas, J. S., & Loring, B. A. (2008). *Moderating usability tests [electronic resource] : principles and practices for interacting*. Retrieved from http://www.loc.gov/catdir/enhancements/fy0827/2007046074-d.html. Retrieved from e-Book ScienceDirect database.

Ewing, J., & Miller, D. (2002). A framework for evaluating computer supported collaborative learning. *Journal of Educational Technology & Society, 5* (1 (Special Issue on "Integrating Technology into Learning and Working (Part 1)")), 112-118 Retrieved from http://www.ifets.info/journals/5_1/ewing.pdf

Field, A. (2005). *Discovering statistics using SPSS : (and sex, drugs and rock'n'roll)* (2nd ed.). London: SAGE Publications.

Field, A. (2006). *Research Methods II: Reliability Analysis.* Retrieved 3 May, 2010, from http://www.statisticshell.com/reliability.pdf

Folmer, E., & Bosch, J. (2003). Usability Patterns in Software Architecture. Retrieved from http://is.ls.fi.upm.es/status/results/patterns.pdf

ForakerDesign. (2005). *Your Online Guide to Usability Resources - Methods*. Retrieved 16 Jul. 2008, from http://www.usabilityfirst.com/methods/index.txl

Frøkjær, E., Hertzum, M., & Hornbæk, K. (2000). *Measuring usability: are effectiveness, efficiency, and satisfaction really correlated?* Paper presented at the meeting of the Proceedings of the SIGCHI conference on Human factors in computing systems, The Hague, The Netherlands. doi:http://doi.acm.org/10.1145/332040.332455

Garson, G. D. (2010). *Reliability Analysis*. Retrieved 13 May, 2010, from http://faculty.chass.ncsu.edu/garson/PA765/reliab.htm

Gellner, M., & Forbrig, P. (2003). Extreme Evaluations – Lightweight Evaluations for Software Developers. Retrieved from http://www.se-hci.org/bridging/interact/Gellner.pdf

Gliem, J. A., & Gliem, R. R. (2003). *Calculating, Interpreting, and Reporting Cronbachs Alpha Reliability Coefficient for Likert-Type Scales*. Paper presented at the meeting of the 2003 Midwest Research to Practice Conference in Adult, Continuing, and Community Education, Retrieved from https://scholarworks.iupui.edu/bitstream/handle/1805/344/Gliem+&+Gliem.pdf?sequence=1

Gomoll, K., & Nicol, A. (1990). *Discussion of guidelines for user observation*. Retrieved 10 Apr, 2009, from http://grouplab.cpsc.ucalgary.ca/saul/hci_topics/assignments/usability/ass2_usabil.html

González, M. P., Collazos, C. A., & Granollers, T. (2006). Guidelines and Usability Principles to Design and Test Shared-Knowledge Awareness for a CSCL Interface In *Groupware: Design, Implementation, and Use* (Vol. 4154, pp. 102-117): Springer Berlin / Heidelberg. Retrieved from http://www.springerlink.com/content/4n72362151538851/. Retrieved 17 Jan 2009. doi:10.1007/11853862

Graham, C. R., & Misanchuk, M. (2004). Computer-Mediated Learning Groups: Benefits and Challenges to Using Groupwork in Online Learning Environments. In T. S. Roberts (Ed.), *Online collaborative learning [electronic resource] : theory and practice* (pp. 181-202). Hershey PA: Idea Group Publishing. Retrieved 8 Jan 2009

Green, D., & Pearson, J. M. (2006). Development of a Web Site Usability Instrument based on ISO 9241-11. *Journal of Computer Information Systems, 67-72*(1), 66-72.

Greenberg, S., Fitzpatrick, G., Gutwin, C., & Kaplan, S. (2000). Adapting the Locales Framework for Heuristic Evaluation of Groupware. *Australian Journal of Information Systems (AJIS), 7(2)*, 102-108.

Gutwin, C., & Greenberg, S. (1999). The effects of workspace awareness support on the usability of real-time distributed groupware. *ACM Trans. Comput.-Hum. Interact., 6*(3), 243-281. doi:http://doi.acm.org/10.1145/329693.329696

Gutwin, C., & Greenberg, S. (2000). The Mechanics of Collaboration: Developing Low Cost Usability Evaluation Methods for Shared Workspaces*IEEE.* Symposium conducted at the meeting of the Enabling Technologies: Infrastructure for Collaborative Enterprises, 2000. (WET ICE 2000). Proeedings. IEEE 9th International Workshops on, New York. Retrieved from http://ieeexplore.ieee.org.ezproxy.auckland.ac.nz/stamp/stamp.jsp?tp=&arnumber=883711&isnumber=19118 doi:10.1109/ENABL.2000.883711

Hancock, D. R., & Algozzine, B. (2006). Doing case study research [electronic resource] : a practical guide for beginning researchers. In. New York: Teachers College Press.

Henry, S. L. (2008). *Web Content Accessibility Guidelines (WCAG) Overview*. Retrieved 22 Jun., 2009, from http://www.w3.org/WAI/intro/wcag.php

Hollingsed, T., & Novick, D., G. (2007). *Usability inspection methods after 15 years of research and practice*. Paper presented at the meeting of the Proceedings of the 25th annual ACM international conference on Design of communication, El Paso, Texas, USA. Retrieved from http://delivery.acm.org.ezproxy.auckland.ac.nz/10.1145/1300000/1297200/p249-hollingsed.pdf?key1=1297200&key2=3997076221&coll=ACM&dl=ACM&CFID=70804985&CFTOKEN=48408330 doi:http://doi.acm.org/10.1145/1297144.1297200

Holz, H., J., Applin, A., Haberman, B., Joyce, D., Purchase, H., & Reed, C. (2006). Research methods in computing: what are they, and how should we teach them? *SIGCSE Bull., 38*(4), 96-114. doi:http://doi.acm.org/10.1145/1189136.1189180

Hom, J. *Thinking Aloud Protocol* Retrieved 20 June, 2010, from http://usability.jameshom.com/

Hu, G., & Chang, K. H. (2006). *Web sites usability, usability requirements specification & usability evaluation*. Paper presented at the meeting of the Proceedings of the 44th annual Southeast regional conference, Melbourne, Florida. Retrieved from http://portal.acm.org.ezproxy.auckland.ac.nz/ft_gateway.cfm?id=1185640&type=pdf&coll=ACM&dl=ACM&CFID=7825550&CFTOKEN=29169951 doi:http://doi.acm.org/10.1145/1185448.1185640

Hwang, W., & Salvendy, G. (2010). Number of People Required for Usability Evaluation: The 10±2 Rule. *Communications of the ACM, 53*(5), 130-133. doi:10.1145/1735223.1735255

Ingram, A. L., & Hathorn, L. G. (2004). Methods for Analyzing Collaboration in Online Communications In T. S. Roberts (Ed.), *Online collaborative learning: theory and practice* (pp. 215-241). Hershey PA: Idea Group Publishing. Retrieved 10 Jan 2009

Intraclass correlation. Retrieved 3 May, 2010, from http://en.wikipedia.org/wiki/Intra-class_correlation_coefficient

Ivory, M. Y., & Hearst, M. A. (2001). *The state of the art in automating usability evaluation of user interfaces*. Paper presented at the meeting of the ACM Computing Surveys, Retrieved from http://ezproxy.auckland.ac.nz/login?url=http://proquest.umi.com/pqdweb?did=109542419&Fmt=7&clientId=13395&RQT=309&VName=PQD

Karoulis, A., & Pombortsis, A. (2003). Heuristic Evaluation of Web-based ODL Programs. In C. Ghaoui (Ed.), *Usability Evaluation of Online Learning Programs* (pp. 88-109). Hershey PA, USA: Information Science publishing. Retrieved 16 Jul. 2008

Kildare, R., Williams, R. N., & Hartnett, J. (2006). An online tool for learning collaboration and learning while collaborating*Australian Computer Society, Inc.* Symposium conducted at the meeting of the Proceedings of the 8th Austalian conference on Computing education - Volume 52, Hobart, Australia. Retrieved from http://delivery.acm.org.ezproxy.auckland.ac.nz/10.1145/1160000/1151883/p101-kildare.pdf?key1=1151883&key2=2869029221&coll=ACM&dl=ACM&CFID=15076723&CFTOKEN=17043587

Kirakowski, J. (2000). *Questionnaires in Usability Engineering: A List of Frequently Asked Questions (3rd Ed.)*. Retrieved 20 June, 2010, from http://www.ucc.ie/hfrg/resources/qfaq1.html

Konstantinidis, A., Tsiatsos, T., & Pomportsis, A. (2009). Collaborative virtual learning environments: design and evaluation. *Multimedia Tools & Applications, 44*(2), 279-304.

Koohang, A. (2004). A Study of Users' Perceptions Toward E-Learning Courseware Usability. *International Journal on E-Learning, 3*(2), 10-17. Retrieved from http://www.editlib.org.ezproxy.auckland.ac.nz/p/12792

Kurfess, F. J. (2005). *User-Centered Design and Development.* Retrieved 1 June, 2008, from http://users.csc.calpoly.edu/~fkurfess/Courses/484/W06/Slides/Chapter11ID-FJK.ppt

Lani, J. (2009). *Dissertation Statistics Help: Cronbach's Alpha Rule of Thumb*. Retrieved 3 May, 2010, from http://statisticssolutions.blogspot.com/2009/03/cronbachs-alpha-rule-of-thumb.html

LeBreton, J. M., & Senter, J. L. (2008). Answers to 20 Questions About Interrater Reliability and Interrater Agreement. *Organizational Research Methods, 11*(4), 815-852. Retrieved from http://orm.sagepub.com/cgi/reprint/11/4/815 doi:10.1177/1094428106296642

Leech, N. L., Barrett, K. C., & Morgan, G. A. (2005). *SPSS for Intermediate Statistics: Use and Interpretation* (2nd ed.). Retrieved from http://www.netlibrary.com/Details.aspx

Lewis, J. R. (1995). IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use. Technical Report 54.786. *International Journal of Human-Computer Interaction, 7*(1), 41.

Li, Q., Lau, R. W. H., Shih, T. K., & Li, F. W. B. (2008). Technology supports for distributed and collaborative learning over the internet. *ACM Transactions on Internet Technology, 8*(2), 1-24. doi:http://doi.acm.org/10.1145/1323651.1323656

López-Jaquero, V., Montero, F., Fernández-Caballero, A., & Lozano, M. D. (2003). Usability Metrics in Adaptive Agent-based Tutoring Systems. Retrieved from http://www.info-ab.uclm.es/personal/AntonioFdez/download/papers/conference/HCII2003-usability.pdf

Lund, K. (2004). Human Support in CSCL: What, for Whom, and by Whom? In J.-W. Strijbos, P. Kirschner & R. Martens (Eds.), *What We Know About CSCL* (Vol. 3, pp. 167-198): Springer Netherlands. Retrieved from http://dx.doi.org/10.1007/1-4020-7921-4_7. doi:10.1007/1-4020-7921-4_7

Mack, R. L., & Nielsen, J. (1994). Executive Summary. In J. Nielsen & R. L. Mack (Eds.), *Usability inspection methods* (pp. 1-23). New York: John Wiley & Sons, Inc. Retrieved 14 Jul. 2008

Mandryk, R. L., & Inkpen, K. M. (2004). *Physiological indicators for the evaluation of co-located collaborative play*. Paper presented at the meeting of the Proceedings of the 2004 ACM conference on Computer supported cooperative work, Chicago, Illinois, USA. Retrieved from http://delivery.acm.org.ezproxy.auckland.ac.nz/10.1145/1040000/1031625/p102-mandryk.pdf?key1=1031625&key2=3314673221&coll=ACM&dl=ACM&CFID=12025639&CFTOKEN=38359002 doi:http://doi.acm.org/10.1145/1031607.1031625

Marghescu, D. (2009). Usability Evaluation of Information Systems: A Review of Five International Standards In *Information Systems Development: Challenges in Practice, Theory, and Education* (Vol. 1 pp. 131-142): Springer US. doi:10.1007/978-0-387-68772-8

Matera, M., Rizzo, F., & Carughi, G. T. (2006). Web Usability: Principles and Evaluation Methods. *Web Engineering*, 143-180. Retrieved from http://www.springerlink.com.ezproxy.auckland.ac.nz/content/hp5t836gxxv33m47/fulltext.pdf doi:10.1007/3-540-28218-1

McGee, M. (2004). *Master usability scaling: magnitude estimation and master scaling applied to usability measurement*. Paper presented at the meeting of the Proceedings of the SIGCHI conference on Human factors in computing systems, Vienna, Austria. doi:http://doi.acm.org/10.1145/985692.985735

McGraw, K. O., & Wong, S. P. (1996). Forming Inferences About Some Intraclass Correlation Coefficients. *Psychological Methods, 1*(1), 30-46

McNeill, P., & Chapman, S. (2005). *Research methods* (3rd ed.). New York, NY: Routledge.

Melzer, I., Shtilman, I., Rosenblatt, N., & Oddsson, L. I. (2007). Reliability of voluntary step execution behavior under single and dual task conditions. *Journal of NeuroEngineering and Rehabilitation, 4*(16). doi:10.1186/1743-0003-4-16

Merriam, S. B. (2009). *Qualitative research: a guide to design and implementation*. San Francisco, Calif.: Jossey-Bass.

Molich, R., & Dumas, J. S. (2008). Comparative usability evaluation (CUE-4). *Behaviour & Information Technology, 27*(3), 263 - 281.

Muthayan, S. (2003). Open access research and the public domain in South African universities: The Public Knowledge Project's Open Journal Systems. Symposium conducted at the meeting of the International Symposium on Open Access and the Public Domain in Digital Data and Information for Science, UNESCO, Paris. Retrieved from http://pkp.sfu.ca/files/Muthayan.pdf

Nakamichi, N., Shima, K., Sakai, M., & Matsumoto, K.-i. (2006). *Detecting low usability web pages using quantitative data of users' behavior*. Paper presented at the meeting of the Proceedings of the 28th international conference on Software engineering, Shanghai, China. Retrieved from http://delivery.acm.org.ezproxy.auckland.ac.nz/10.1145/1140000/1134365/p569-nakamichi.pdf?key1=1134365&key2=7617894221&coll=ACM&dl=ACM&CFID=7825550&CFTOKEN=29169951 doi:http://doi.acm.org/10.1145/1134285.1134365

Neale, D. C., & Carroll, J., M. (1999). *Multi-faceted evaluation for complex, distributed activities*. Paper presented at the meeting of the Proceedings of the 1999 conference on Computer support for collaborative learning, Palo Alto, California. Retrieved from http://delivery.acm.org.ezproxy.auckland.ac.nz/10.1145/1160000/1150293/a53-neale.pdf?key1=1150293&key2=2075773221&coll=GUIDE&dl=GUIDE&CFID=5922550&CFTOKEN=73015519

Neale, D. C., Carroll, J. M., & Rosson, M. B. (2004). *Evaluating computer-supported cooperative work: models and frameworks*. Paper presented at the meeting of the Proceedings of the 2004 ACM conference on Computer supported cooperative work, Chicago, Illinois, USA. Retrieved from http://portal.acm.org.ezproxy.auckland.ac.nz/ft_gateway.cfm?id=1031626&type=pdf&coll=ACM&dl=ACM&CFID=12025639&CFTOKEN=38359002 doi:http://doi.acm.org/10.1145/1031607.1031626

Nichols, D. P. (1998). *Choosing an Intraclass Correlation Coefficient*. Retrieved 15 May, 2010, from http://www.ats.ucla.edu/stat/spss/library/whichicc.htm

Nichols, D. P. (1999). *My Coefficient a is Negative!* Retrieved 3 May, 2010, from http://www.ats.ucla.edu/stat/spss/library/negalpha.htm

Nielsen, J. (1994). Heuristic Evaluation. In J. Nielsen & R. L. Mack (Eds.), *Usability inspection methods* (pp. 25-62). New York: John Wiley & Sons, Inc. Retrieved 14 Jul. 2008

Nielsen, J. (2005). *Ten Usability Heuristics*. Retrieved 31 Oct 2008, from http://www.useit.com/papers/heuristic/heuristic_list.html

Nielsen, J., & Loranger, H. (2006). *Prioritizing Web usability*. Berkeley, California: New Riders.

NIST. (2007).NISTIR 7432: Common Industry Specification for Usability - Requirements.

Nogueira, J. L., & Garcia, A. C. B. (2003). *Understanding the tradeoffs of Interface Evaluation Methods*. Retrieved 4 Oct, 2004, from http://www.addlabs.uff.br/equipe/cristina/papers/ac1_45.zip

Pfister, H.-R., Wessner, M., Holmer, T., & Steinmetz, R. (1999). Evaluating Distributed Computer-Supported Cooperative Learning (DCSCL): A Framework and Some Data. Retrieved from http://www.iam.unibe.ch/~rvs/research/nlt/talks/pfister.pdf

Pinelle, D., & Gutwin, C. (2000). A review of groupware evaluations Symposium conducted at the meeting of the Enabling Technologies: Infrastructure for Collaborative Enterprises, 2000. (WET ICE 2000). Proeedings. IEEE 9th International Workshops on Retrieved from http://ieeexplore.ieee.org.ezproxy.auckland.ac.nz/stamp/stamp.jsp?arnumber=883709&isnumber=19118

Pinelle, D., & Gutwin, C. (2002). *Groupware walkthrough: adding context to groupware usability evaluation*. Paper presented at the meeting of the Proceedings of the SIGCHI conference on Human factors in computing systems: Changing our world, changing ourselves, Minneapolis, Minnesota, USA. Retrieved from http://delivery.acm.org.ezproxy.auckland.ac.nz/10.1145/510000/503458/p455-pinelle.pdf?key1=503458&key2=7825292321&coll=ACM&dl=ACM&CFID=19527122&CFTOKEN=75144380 doi:http://doi.acm.org/10.1145/503376.503458.

Pinelle, D., & Gutwin, C. (2008). Evaluating teamwork support in tabletop groupware applications using collaboration usability analysis. *Personal and Ubiquitous Computing. Special Issue: User-centred design and evaluation of ubiquitous groupware, 12*(3), 237-254. doi:http://dx.doi.org/10.1007/s00779-007-0145-4

Pinelle, D., Gutwin, C., & Greenberg, S. (2003). Task analysis for groupware usability evaluation: Modeling shared-workspace tasks with the mechanics of collaboration. *ACM Trans. Comput.-Hum. Interact., 10*(4), 281-311. doi:http://doi.acm.org/10.1145/966930.966932

PKP. (2008a). OJS in an Hour. An Introduction to Open Journal Systems. Version 2.2.1.0. Retrieved from http://pkp.sfu.ca/files/OJSinanHour.pdf

PKP. (2008b). *Open Journal System*. Retrieved 30 April, 2008, from http://pkp.sfu.ca/?q=ojs

PKP. (2010). *OJS Documentation*. Retrieved 6 Jun 2010, from http://pkp.sfu.ca/ojs_documentation

Preece, J. (2000). *Online communities: designing usability, supporting sociability*. New York: John Wiley.

Quesenbery, W. (2003). The Five Dimensions of Usability. In M. J. Albers & B. Mazur (Eds.), *Content & Complexity: Information Design in Technical Communication* (pp. 81-102). Mahwah, New Jersey: Lawrence Erlbaum Associates. Retrieved 9 Nov 2008

Ross, S., Ramage, M., & Rogers, Y. (1995). PETRA: Participatory Evaluation Through Redesign and Analysis. *Cognitive Science Research Paper (CSRP) No. 375*. Retrieved from http://www.comp.lancs.ac.uk/computing/research/cseg/projects/evaluation/lit_methods.html

Rubin, J. (1994). *Handbook of usability testing : how to plan, design, and conduct effective tests* New York: Wiley.

Rubin, J., & Chisnell, D. (2008). *Handbook of usability testing: how to plan, design, and conduct effective tests* (2nd ed.). Indianapolis, Indiana: Wiley Publishing, Inc.

Ryu, H. (2007). Walkthroughs in Web Usability: Cognitive, Activity, and Heuristic Walkthrough. In P. Zaphiris & S. Kurniawan (Eds.), *Human Computer Interaction Research in Web Design and Evaluation.* (pp. 229-256). London: Idea Group Publishing.

Santos, J. R. A. (1999). *Cronbach's Alpha: A Tool for Assessing the Reliability of Scales*. Retrieved 3 May, 2010, from http://www.joe.org/joe/1999april/tt3.php

Sauro, J., & Kindlund, E. (2005). *A method to standardize usability metrics into a single score*. Paper presented at the meeting of the Proceedings of the SIGCHI conference on Human factors in computing systems, Portland, Oregon, USA. Retrieved from http://delivery.acm.org.ezproxy.auckland.ac.nz/10.1145/1060000/1055028/p401-sauro.pdf?key1=1055028&key2=9659134221&coll=ACM&dl=ACM&CFID=6987341&CFTOKEN=23425010 doi:http://doi.acm.org/10.1145/1054972.1055028

Schmettow, M., & Vietze, W. (2008). *Introducing item response theory for measuring usability inspection processes*. Paper presented at the meeting of the Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems: Measuring, Business, and Voting, Florence, Italy. Retrieved from http://delivery.acm.org.ezproxy.auckland.ac.nz/10.1145/1360000/1357196/p893-schmettow.pdf?key1=1357196&key2=6293094221&coll=ACM&dl=ACM&CFID=7723426&CFTOKEN=30141889 doi:http://doi.acm.org/10.1145/1357054.1357196

Scholtz, J. (2004). *Usability Evaluation*. Retrieved 1 Jun., 2008, from http://www.itl.nist.gov/iad/IADpapers/2004/Usability%20Evaluation_rev1.pdf

Sheard, J., & Markham, S. (2005). Web-based learning environments: developing a framework for evaluation. *Assessment & Evaluation in Higher Education, 30*(4), 353-368.

Shneiderman, B., & Plaisant, C. (2004). *Designing the user interface: Strategies for effective human-computer interaction / Ben Shneiderman, Catherine Plaisant.* Boston: Pearson/Addison Wesley.

Shuttleworth, M. (2009). *Interrater Reliability*. Retrieved 15 May, 2010, from http://www.experiment-resources.com/interrater-reliability.html

Skov, M., B., & Stage, J. (2005). *Supporting problem identification in usability evaluations*. Paper presented at the meeting of the Proceedings of the 19th conference of the computer-human interaction special interest group (CHISIG) of Australia on Computer-human interaction: citizens online: considerations for today and the future, Canberra, Australia. Retrieved from ACM Digital Library database.

Smith-Atakan, S. (2006). *Human-computer interaction*. Australia Thomson.

Snyder, C. (2004). Introduction to Usability Test Facilitation. In *Paper Prototyping: The Fast and Easy Way to Design and Refine User Interfaces.* (pp. 171-195). Burlington: Morgan Kaufmann. Retrieved from http://www.sciencedirect.com/science/article/B8713-4P18KJW-V/2/5844209811544aa46037408c65f6555e. Retrieved 10 Apr 2010. doi:10.1016/B978-155860870-2/50035-8

Spada, H., Meier, A., Rummel, N., & Hauser, S. (2005). *A new method to assess the quality of collaborative process in CSCL*. Paper presented at the meeting of the Proceedings of th 2005 conference on Computer support for collaborative learning: learning 2005: the next 10 years!, Taipei, Taiwan. Retrieved from http://portal.acm.org.ezproxy.auckland.ac.nz/citation.cfm?id=1149293.1149375&coll=ACM&dl=ACM&CFID=12025639&CFTOKEN=38359002#

Spool, J. M. (2009). *Moderating with Multiple Personalities: 3 Roles for Facilitating Usability Tests*. Retrieved 5 May, 2010, from http://www.uie.com/articles/moderating_multiple_personalities/

Ssemugabi, S., & Villiers, R. D (2007a). Usability and Learning: A Framework for Evaluation of Web-Based e-Learning Applications. In C. Montgomerie & J. Seale (Chair), *AACE.* Symposium conducted at the meeting of the World Conference on Educational Multimedia, Hypermedia and Telecommunications 2007, Vancouver, Canada.

Ssemugabi, S., & Villiers, R. D. (2007b). *A comparative study of two usability evaluation methods using a web-based e-learning application*. Paper presented at the meeting of the Proceedings of the 2007 annual research conference of the South African institute of computer scientists and information technologists on IT research in developing countries, Port Elizabeth, South Africa. Retrieved from http://portal.acm.org.ezproxy.auckland.ac.nz/ft_gateway.cfm?id=1292507&type=pdf&coll=ACM&dl=ACM&CFID=7908696&CFTOKEN=54401013 doi:http://doi.acm.org/10.1145/1292491.1292507

Steves, M. P., Morse, E., Gutwin, C., & Greenberg, S. (2001). *A comparison of usage evaluation and inspection methods for assessing groupware usability*. Paper presented at the meeting of the Proceedings of the 2001 International ACM SIGGROUP Conference on Supporting Group Work, Boulder, Colorado, USA. Retrieved from http://portal.acm.org.ezproxy.auckland.ac.nz/ft_gateway.cfm?id=500306&type=pdf&coll=ACM&dl=ACM&CFID=7843053&CFTOKEN=83891791 doi:http://doi.acm.org/10.1145/500286.500306

Suddaby, R. (2006). From the editors: What grounded theory is not. *The Academy of Management Journal Archive, 49*(4), 633-642.

Trochim, W. M. K. (2006). *Reliability*. Retrieved 3 May, 2010, from http://www.socialresearchmethods.net/kb/reliable.php

Tselios, N., Avouris, N., & Komis, V. (2008). The effective combination of hybrid usability methods in evaluating educational applications of ICT: Issues and challenges [Article]. *Education & Information Technologies, 13*(1), 55-76. doi:10.1007/s10639-007-9045-5

UCLA. *SPSS FAQ: What does Cronbach's alpha mean?* Retrieved 3 May, 2010, from http://www.ats.ucla.edu/stat/spss/faq/alpha.html

UsabilityNet. (2006). *International standards for HCI and usability*. Retrieved 14 Oct 2008, from http://www.usabilitynet.org/tools/r_international.htm

W3C. (2008). *WEB ACCESSIBILITY QUICKTIPS: WCAG 2.0 at a Glance*. Retrieved 13 Jun, 2009, from http://www.w3.org/WAI/WCAG20/glance/WCAG2-at-a-Glance-a4.pdf

Wells, M. A., & Brook, P. W. (2004). Conversational KM Student Driven Learning*Australian Computer Society, Inc. Darlinghurst, Australia.* Symposium conducted at the meeting of the Proceedings of the sixth conference on Australasian computing education, Dunedin, New Zealand.

Wikipedia.org. (2008a). *Open Journal System*. Retrieved 30 Aug 2008, from http://en.wikipedia.org/wiki/Open_Journal_Systems

Wikipedia.org. (2008b). *Public Knowledge Project*. Retrieved 30 Aug 2008, from http://en.wikipedia.org/wiki/Public_Knowledge_Project

Willinsky, J. (2005). Open Journal Systems: An example of open source software for journal management and publishing. *Library Hi Tech, 23*(4), 504 - 519. doi:10.1108/07378830510636300

Wolz, U., Palme, J., Anderson, P., Chen, Z., Dunne, J., Karlsson, G., et al. (1997). Computer-mediated communication in collaborative educational settings (report of the ITiCSE '97 working group on CMC in collaborative educational settings)*ACM New York, NY, USA.* Symposium conducted at the meeting of the The supplemental proceedings of the conference on Integrating technology into computer science education: working group reports and supplemental proceedings, Uppsala, Sweden. Retrieved from http://delivery.acm.org.ezproxy.auckland.ac.nz/10.1145/270000/266100/p51-wolz.pdf?key1=266100&key2=7965249221&coll=ACM&dl=ACM&CFID=14775664&CFTOKEN=99421142 doi:http://doi.acm.org/10.1145/266057.266100

Xerox-Corporation. (1995). *Heuristic Evalualtion - A System Checklist*. Retrieved 13 Feb, 2009, from http://www.stcsig.org/usability/resources/toolkit/he_cklst.doc

Yin, R. K. (1994). *Case Study Research: Design and Methods* (2nd ed.). Thousand Oaks, CA: SAGE Publications.

Yin, R. K. (2003a). *Case Study Research: Design and Methods* (3rd ed.). Thousand Oaks: CA: SAGE Publications.

Yin, R. K. (2003b). *Applications of Case study Research* (2nd ed.). Thousand Oaks, California: Sage Publications.

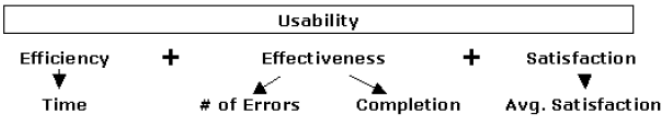
Yin, R. K. (2004). *Case Study Methods (revised draft)*. Retrieved 29 Mar 2008, from http://029c7c0.netsolhost.com/Docs/AERAdraft.pdf

Zhang, Z. (2007). Usability Evaluation. In P. Zaphiris & S. Kurniawan (Eds.), *Human Computer Interaction Research in Web Design and Evaluation* (pp. pp209- 228). London: Idea Group Publishing. Retrieved 15 Oct 2008

Zhang, Z. (2008). *Usability Evaluation Methods*. Retrieved 1 Jun. 2008, from http://www.usabilityhome.com/

Appendixes

Appendix 1 List of some frameworks for UE

#	Frameworks (attributes /criteria)	Number of attributes /criteria (Context)	References
1	Learnability, efficiency, memorability, errors, and user satisfaction. Depending on the type of application one attribute might be more critical than another	5 (Usability engineering)	(Scholtz, 2004)
2	Ease of use, learnability, memorability, adjustability individual needs, task accomplishment, and overall users' satisfaction	6 (Interface evaluation)	(Nogueira & Garcia, 2003)
3	1) *Effectiveness,* e.g. quality of solution and error rates. 2) *Efficiency,* e.g. task completion time and learning time. 3) *Satisfaction,* e.g. attitude rating scales e.g. SUMI. efficiency, effectiveness, and satisfaction should be considered independent aspects of usability, unless domain specific studies suggest otherwise.	3 (computer systems in general)	(Frøkjær, et al., 2000)
4	*1) Effectiveness:* how well a method discovers usability problems, *2) Ease of use:* how easy a method is to employ, *3) Ease of learning:* how easy a method is to learn, and 4) *Applicability:* how widely applicable a method is to WIMP (window, icon, menu, pointing device) and/or Web UIs than to those originally applied.	4 (user interface design)	(Ivory & Hearst, 2001)
5	*Usefulness* concerns the degree to which a product enables a user to achieve his or her goals, and it's an assessment of the user's willingness to use the product at all. Without that motivation, other measures make no sense, because the product will just sit on the shelf. *Efficiency* is the quickness with which the user's goal can be accomplished accurately and completely and is usually a measure of time. *Effectiveness* refers to the extent to which the product behaves in the way that users expect it to and the ease which users can use it to do what they intend. This is usually measured quantitatively with error rate. *Learnability* is part of effectiveness and has to do with the user's ability to operate the system to some defined level of competence after some predetermined amount and period of training. It can also refer to the ability of infrequent users to relearn the system after periods of inactivity. *Satisfaction* refers to the user's perceptions, feelings, and opinions of the product, usually captured through both written and oral questioning. Users are more likely to perform well on the product that meets their needs and provides satisfaction than one that does not. *Accessibility* is about having access to the products needed to accomplish a goal.	6 (usability testing)	(Rubin & Chisnell, 2008)
6	***Summated Usability Metric (SUM)*** four dimensions—task completion, error counts, task time, and satisfaction scores  Quantitative Model of Usability	4 (Quantitative Model of Usability to standardize usability metrics into a single score)	(Sauro & Kindlund, 2005)

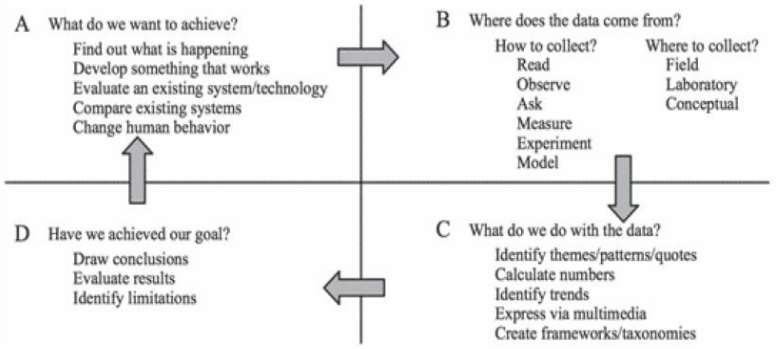
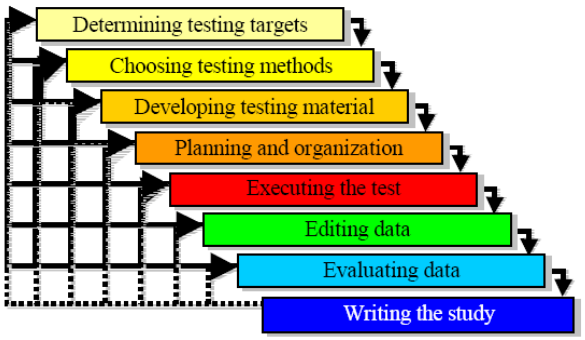
7	***more directly to practical evaluation*** 1) *Time to learn.* How long does it take for typical members of the user community to learn how to use the actions relevant to a set of tasks? 2) *Speed of performance*. How long does it take to carry out the benchmark tasks? 3) *Rate of errors by users.* error handling is a critical component of interface usage. 4) *Retention over time.* Retention may be linked closely to time to learn, and frequency of use plays an important role. 5) *Subjective satisfaction.* use interview or written surveys that include satisfaction scales and space for free-form comments.	5 (user interface)	(Shneiderman & Plaisant, 2004)
8	***5E dimensions of Usability*** 1) *Effective*: the completeness and accuracy with which users achieve their goals. 2) *Efficient*: the speed (with accuracy) with which users can complete their tasks. Resources include the number of individual actions a user must take and the time spent on them. 3) *Engaging:* How well the interface draws the user into the interaction and how pleasant and satisfying it is to use 4) *Error tolerant*: how well the design prevents errors, or helps with recovery from those that do occur. 5) *Easy to Learn*: good instructions, prompts, examples, or hints can provide enough information to create an interface in which the user can extend prior experiences into a new context. an easy to learn interface is both consistent and predictable.	5 (User-centred design)	(González, et al., 2006; Mandryk & Inkpen, 2004; Quesenbery, 2003)
9	***Nine dimensions for Evaluation of the quality of collaborative process*** It is a method for assessing the quality of collaborative process in computer-supported problem-solving and learning settings and on a relatively global level, resulting in quantitative ratings that can be subjected to statistical analyses. The method consists of nine dimensions: *1)* *sustaining mutual understanding* and *coordinating communication* refer to basic communication processes which form a prerequisite for successful collaboration; *2)* *information pooling* and *reaching consensus* are relevant for the construction and maintenance of a shared understanding; *3)* *Task division, time management* and *technical coordination* are three dimensions reflecting the coordination of collaborative activities; 4) *and shared task alignment* and *sustaining commitment* are the motivational aspects.	9 (assessing collaborative process, quantitatively rating nine qualitatively defined characteristic dimensions of collaboration)	(Spada, et al., 2005)
10	***Neilsen's ten heuristics (rules) (2001) (web usability)*** 1) Visibility of system status 2) Match between system and the real world. 3) User control and freedom. 4) Consistency and standards. 5) Error prevention. 6) Helping users recognise, diagnose and recover from errors. 7) Recognition rather than recall. 8) Flexibility and efficiency of use. 9) Aesthetic and minimalist design. 10) Help and documentation.	10 (Heuristics User Interface Design)	(Karoulis & Pombortsis, 2003; Nielsen, 2005; Quesenbery, 2003)
11	***Heuristics of web usability evaluation -An extended version of the original heuristics*** 1) Visibility of current web page status 2) Match between the system world and the real world based on targeted users	10 (Heuristics of web usability)	(Nielsen, 1994; Ryu, 2007)

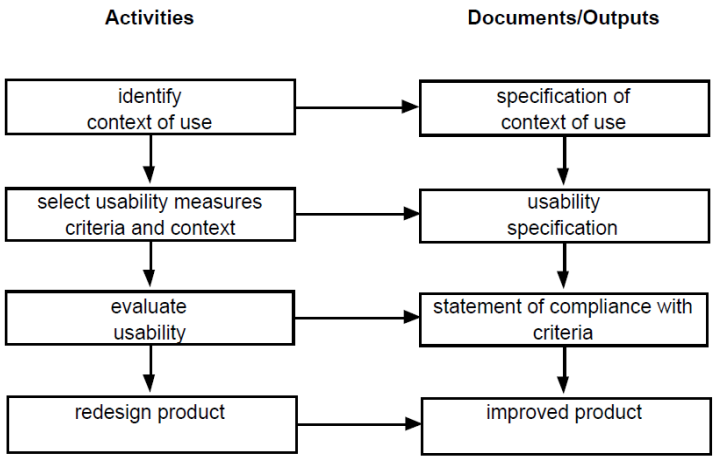
	3) Support user control to web navigation and relevant link 4) Consistent web design and conformation to standards 5) Error prevention with informative contents 6) Recognition rather than recall 7) Flexibility and efficiency of use for frequent visitors 8) Aesthetic and minimal scrolling design 9) Help users recognise, diagnose, and recover from errors 10) Help and documentation		
12	***Web Content Accessibility Guidelines WCAG 2.0*** 1) *Perceivable:* "Make Content perceivable by any users." 2) *Understandable*: "make content and controls understandable to as many users as possible." 3) *Robust*: "Use web technologies that maximize the ability of the content to work with current and future accessibility technologies and user agents." 4) *Operable*: "Ensure that interface elements in the content are operable by any user."	4 (Web Content Accessibility)	(Abascal, et al., 2007; W3C, 2008)
13	***Holistic Usability Model*** *for distributed simulation systems* The main groupings for this holistic usability model are designing for different users, usefulness, ease of use, satisfaction, system context, and user experience. 55 criteria can be grouped to eight types of measures. 1) End user Needs and goals 2) End user interface(s) 3) Control features 4) Data visualisation and analysis 5) Programming 6) Installation 7) Training 8) Documentation	55 See Figure 2-6 on pg 19 (distributed simulation systems)	(Dawson, 2006)
14	***The framework comprising usability attributes, properties and patterns:*** (as a source to inform architecture design for usability) 1) *Learnability* - how quickly and easily users can begin to do productive work with a system that is new to them, combined with the ease of remembering the way a system must be operated. 2) *Efficiency of use* - the number of tasks per unit time that the user can perform using the system. 3) *Reliability in use* - this attribute refers to the error rate in using the system and the time it takes to recover from errors. 4) *Satisfaction* - the subjective opinions that users form in using the system. The following properties have been identified: 1) *Providing feedback* - the system provides continuous feedback as to system operation to the user. 2) *Error management* - includes error prevention and recovery. 3) *Consistency* - consistency of both the user interface and functional operation of the system. 4) *Guidance* - on-line guidance as to the operation of the system. 5) *Minimise cognitive load* - system design should recognise human cognitive limitations, short-term memory etc. 6) *Natural mapping* - includes predictability of operation, semiotic significance of symbols and ease of navigation. 7) *Accessibility* - includes multi-mode access, internationalisation and support for disabled users.	11 (software architecture design - usability)	(Folmer & Bosch, 2003)
15	***Criteria for online journal system evaluation*** **Submission, peer review management, and administrative functions** 1) Support for multiple, discrete publications 2) Multiple administrative roles	28 (online Journal systems)	(Cyzyk & Choudhury, 2008)

	3) Administrative roles configurable 4) Submission into system initiated by authors 5) Editorial workflow configurable per publication 6) Automated email alerts to authors 7) Automated email alerts to editors 8) Automated email alerts to reviewers 9) Stylesheets, customizable look and feel per publication 10) Versioning 11) Archiving **Access, formats, and electronic commerce functions** 12) Accessibility of system 13) Accessibility of document output 14) Internationalization support 15) Output in multiple document formats 16) Document formats supported 17) Plug-in requirements 18) Usability notes 19) Citation linking 20) OpenURL resolver 21) RSS feed 22) Digital rights management 23) Full-text search and retrieval 24) Federated searching 25) Authentication mechanisms 26) Subscription services 27) Electronic commerce functions 28) Context-sensitive Help support		
16	***Five Web Site Usability Dimensions*** 1) Effectiveness: the completeness and accuracy with which users achieve specified goals 2) Efficiency: the speed with which work can be done. 3) Level of engagement: how pleasant, satisfying or interesting an interface is to use. 4) Error tolerance: how well the product prevents errors, and helps the user recover from any errors that do occur 5) Ease of learning: how well the web site supports both initial orientation and deeper learning.	5 (web site usability)	(Green & Pearson, 2006)
17	**Quantitative full-life-cycle method on web evaluation:** A structured use-centred quantitative full-life-cycle method to assist usability requirements specification and usability evaluation of web sites. *Effectiveness Efec:* Effectiveness's basic use features include ***accuracy*** (*Agt*) and ***completeness*** (*Cgt*), and we define: (1) $E_{fec} = 50\% C_{gt} + 50\% A_{gt}$ *Efficiency (E):* the total actual amount of time expended is *T*, the wasted amount of time is *Tw*, we define: (2) $E = \frac{T - T_w}{T}$ ***Goal-task's use-centred usability (U_{gt})***: $U_{gt} = (25\% U_p + 25\% U_{p-\text{int}} + 25\% S + 25\% E) E_{fec}$ *Satisfaction of a use (S); Use interaction process aptness (Up)*;*Use interaction interface and presentation aptness (Up int)*; The quantitative web site usability (*U*) can be defined as: (3) $U = (W_1(\sum W_{gt_i} U_{gt_i}) + W_2 U_{nav}) U_{uni-\text{int}}$ *W1* is the weight of the combined usability of all the normal tasks in a system; *Unav* is the system navigation mechanism aptness and	10 variables quantitative UE for web sites	(Hu & Chang, 2006)

	W2 is its weight, $U_{uni\text{-}int}$ is the universal interface and presentation aptness. *Ugti* is the quantitative usability of the number ***i*** designed goal-task; *Wgti* is the weight of the number ***i*** designed goal-task; So, the system usability *U* is a composite use feature that signifies the comprehensive overall usability in percentage of a system by combining together all its designed goal-tasks' usability and the system navigation mechanism aptness and then taking into account the universal interface and presentation aptness as a critical factor that severely affects the overall usability of a system.		
18	***Evaluation criteria for web-based learning*** – a multi-faceted framework three categories for evaluating WBL applications • Category 1: General interface usability criteria (based on Nielsen's heuristics, modified for e-learning context) 1) Visibility of system status 2) Match between the system and the real world i.e. match between designer model and user model 3) Learner control and freedom 4) Consistency and adherence to standards 5) Error prevention, in particular, prevention of peripheral usability-related errors 6) Recognition rather than recall 7) Flexibility and efficiency of use 8) Aesthetics and minimalism in design 9) Recognition, diagnosis, and recovery from errors 10) Help and documentation • Category 2: Website-specific criteria for educational websites 11) Simplicity of site navigation, organisation and structure 12) Relevance of site content to the learner and the learning process • Category 3: Learner-centred instructional design, grounded in learning theory, aiming for effective learning 13) Clarity of goals, objectives and outcomes 14) Effectiveness of collaborative learning (where such is available) 15) Level of learner control 16) Support for personally significant approaches to learning 17) Cognitive error recognition, diagnosis and recovery 18) Feedback, guidance and assessment 19) Context meaningful to domain and learner *20)* Learner motivation, creativity and active learning	20 (web-based e-learning applications)	(Ssemugabi & Villiers, 2007a)

Appendix 2 List of some methodologies for UE

#	Methodology of UE	Description	References /goal(s) of UE
1	**Computing research Methods (CRM)**	It grounded in four questions and each question anchors a quadrant in the process of computing research.  	(Holz, et al., 2006) Computing research in general
2	**Three-step holistic approach**	**three-step**:(1) the development of a holistic usability framework for a system, (2) surveys of users to validate and refine the framework, and to determine attribute weights, and (3) the application of this framework to the existing system, including the development of a technique to measure holistic usability. Taking a holistic viewpoint is especially helpful with a large, complex system, in that the benefits of improved usability can be multiplied by looking not only at the user interface, but also at various aspects of system design, installation, maintenance, and use.	(Dawson, 2006) Developing a holistic usability framework
3	**Eight Phase Pattern**	eight phased pattern  	(Gellner & Forbrig, 2003) Testing targets e.g. a system
4	**Five overall activities of usability evaluation**	(1) determine basics, (2) plan process, (3) create test situation, (4) conduct test and (5) interpret data	(Rubin, 1994; Rubin & Chisnell, 2008) UE in general
5	**GQM**	**Goals** - the overall aim of the evaluation **Questions** - directing the focus of the evaluation **Metric** - being used by evaluators to quantify answers to questions.	(Preece, 2000) UE in general

6	**Quality Plan**	 The plan includes 4 steps. The difference between usability and the quality of a work system in use is a matter of focus. When usability is evaluated, the focus is on improving a product. The purpose of identifying context of use is to ensure validity of test results the users, tasks and environments used for the evaluation should match the real context of use as closely as possible.	(Bevan, 1995) Evaluating usability - one aspect of a product quality → improving product quality;
7	**Scenario-based evaluation**	Scenario describes what the users are expected to do, such as analyze imagery or create a logistical plan in a given context. It usually also specifies the characteristics of the group that should carry it out, and the social protocols which should be in place. The evaluation involves: 1) Identifying the evaluation goal (comparison between systems, appropriate of system for requirements, etc) and the system(s) to be evaluated; 2) Formulating hypotheses to be tested 3) Designing or selecting appropriate scenarios for the test 4) Identifying the appropriate measures 5) Running subjects through the scenarios while the experimenter takes measurements and observations of the interaction for later analysis. 6) Analyzing the data to verify or disprove the hypotheses.	(Damianos, et al., 1999) Testing hypotheses
8	**DECIDE**	***D*etermine** the *goals* the evaluation addresses. ***E*xplore** the specific *questions* to be answered. ***C*hoose** the *evaluation paradigm* and *techniques* to answer the questions. ***I*dentify** the *practical issues.* ***D*ecide** how to deal with the *ethical issues.* ***E*valuate**, interpret and present the *data.*	(Kurfess, 2005) Identifying practical issues

Appendix 3 Pre-test Questionnaire

Pre-test Questionnaire

Confidentiality Disclaimer: The collected raw information will be de-identified data and may be used in the thesis and subsequent publications. You have the option to review, edit or erase the records.

Participant #: **Date:**

Please answer the following questions. (NA = Not Applicable)

Demography Questions:

1. What is your current professional status?
   ☐ Student ☐ Academic staff ☐ Researcher ☐ other: please state:_____

2. What are your main areas of work/study?
   ☐ Computing
   Please state what they are: e.g. Information System: _____

   ☐ Non-computing
   Please state what they are: e.g. Business: _____

3. Please indicate your age range:
   ☐ ≤ 30 ☐ 30-50 ☐ >50

4. Please indicate your Gender
   ☐ Male ☐ Female

Your experience

5. How long have you been using online systems?
   ☐ ≤ 1yr ☐ 2yrs ☐ 3yrs ☐ 4yrs ☐ > 4yrs ☐ N/A

6. What do/did you use the online systems for?
   ☐ Search information ☐ Shopping
   ☐ Banking ☐ Distance learning or Online study
   ☐ Online Teaching ☐ Other, please state: _____

7. Have you used the **Computer Supported Collaborative Learning** (**CSCL**) systems (e.g. Blackboard)?
   ☐ Yes ☐ No

8. If yes, What **CSCL** systems have you used?
   ☐ Blackboard ☐ Lotus ☐ Cecil ☐ OJS
   ☐ Other, please state: _____

9. How long have you been using the **CSCL** systems?
   ☐ ≤ 1 yr ☐ 1-2 yrs ☐ 2-3 yrs ☐ 3-4 yrs ☐ > 4yrs ☐ N/A

10. Have you used the **CSCL** systems for completing team works / collaborative tasks?
    ☐ Yes ☐ No

11. OJS is a kind of **CSCL** System. Have you worked on this system?
    ☐ Yes ☐ No

12. On the average, how much time do/did you spend per week on **OJS**?
    ☐ Null ☐ ≤ 1hr ☐ 1hr - ≤ 4hrs ☐ 4hrs - ≤ 10hrs ☐ ≥ 10hrs

13. Have you used **OJS** for completing team works / collaborative tasks?
    ☐ Yes ☐ No

14. Please rate your expertise on using **CSCL** systems?
    Novice ⟶ Expert
    ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Approved by the Auckland University of Technology Ethics Committee on 19/May/09. AUTEC Reference number 09/29

Appendix 4 Post–test Questionnaire

Post – Test Questionnaire and Interview questions

Confidentiality Disclaimer: The collected raw information will be de-identified data and may be used in the thesis and subsequent publications. You have the option to review, edit or erase the records.

Date: **Time**: **Participant #:** **Completed Task Titles: Peer Review a submission**

Circle the number which most appropriately reflects your impressions of using the system.

1. **Effectiveness**:	**Strongly Disagree**				**Strongly Agree**	**Not Applicable**
1.1. Completeness						
1.1.1. I am able to complete a task on this system within a proper time frame.	1	2	3	4	5	NA
1.2. *Visibility*						
1.2.1. The system has a good menu or obvious links to support and help me complete a task.	1	2	3	4	5	NA
1.3. Organisation/Design						
1.3.1. T he system interface and design are user friendly and familiar.	1	2	3	4	5	NA
1.3.2. Steps to complete a task follow a logical sequence.	1	2	3	4	5	NA
1.4. Navigability						
1.4.1. It is easy to find where I am and the information I needed when working on a task.	1	2	3	4	5	NA
1.4.2. The information in the system clearly points me to the next step/task in a workflow.	1	2	3	4	5	NA

Please circle a number to rank the importance of each criterion of **Effectiveness** to a system's usability (in general).

	not important at all	**a little important**	**important**	**very important**	**Not Applicable**
Completeness	1	2	3	4	NA
Visibility	1	2	3	4	NA
Organisation/Design	1	2	3	4	NA
Navigability	1	2	3	4	NA

Why do you think the criteria of **Effectiveness** above are important or not important to a system's usability (in general)?

Please write down your comments about **Effectiveness of OJS**: (e.g. what make easy or hard to completing a task/team task? what do you like or dislike? What need to be improved on this system?)

2. Efficiency	**Strongly Disagree**				**Strongly Agree**	**Not Applicable**
2.1. Speed						
2.1.1. I am able to access resources, and work on tasks efficiently.	1	2	3	4	5	NA
2.1.2. The System speed is fast enough.	1	2	3	4	5	NA
2.2. Familiarity/Consistency/Standards						
2.2.1. On a task screen, icons, menus, and information are familiar and understandable to me.	1	2	3	4	5	NA
2.2.2. The layout and interface design are consistent through the whole online system.	1	2	3	4	5	NA
2.3. .*Flexibility/Adaptability/Configurability*						
2.3.1. As a site administrator/editor, I can easily modify/configure forms or templates e.g. email templates provided by the system as necessary.	1	2	3	4	5	NA
2.4. *Effort*						
2.4.1. I don't have to continue remembering information throughout several actions.	1	2	3	4	5	NA
2.4.2. I don't need to learn a lot of things before I can get going with this system.	1	2	3	4	5	NA
2.4.3. I don't need the support of a technical person to be able to use this system.	1	2	3	4	5	NA

Please circle a number to rank the importance of each criterion of **Efficiency** to a system's usability (in general).

	not important at all	**a little important**	**important**	**very important**	**Not Applicable**
Speed	1	2	3	4	NA
Familiarity/Consistency/Standards	1	2	3	4	NA
Flexibility/Adaptability/Configurability	1	2	3	4	NA
Effort	1	2	3	4	NA

Why do you think the criteria of **Efficiency** above are important or not important to a system's usability (in general)?

Please write down your comments about **Efficiency of OJS**: (e.g. what make easy or hard to completing a task/team task? what do you like or dislike? What need to be improved on this system?)

3. **Collaborativity**	**Strongly Disagree**				**Strongly Agree**	**Not Applicable**
3.1. User Management						
3.1.1. It is easy to add /assign users, and manage user roles/accounts on the system.	1	2	3	4	5	NA
3.1.2. Being an Editor, I am able to assign jobs to the teammates.	1	2	3	4	5	NA
3.2. *Awareness*						
3.2.1. After I complete an action, there is an indication that the next action can be started.	1	2	3	4	5	NA
3.3. Communication						
3.3.1. I am able to communicate with the teammates or other users on the system as necessary.	1	2	3	4	5	NA
3.4. User Control/Moderator & Teacher control						
3.4.1. Being a moderator, I am able to give online instructions, and monitor teamwork on the system.	1	2	3	4	5	NA
3.4.2. As a user, I am able to manage my files/notes and the shared files/notes.	1	2	3	4	5	NA
3.5. File/Content Sharing & Management						
3.5.1. Files can be easily uploaded to the system.	1	2	3	4	5	NA
3.5.2. Files can be retrieved easily in the share workspace on the system.	1	2	3	4	5	NA
3.6. Process Tracking/Automated Notification						
3.6.1. After I complete a task, I am able to send a notification to the team.	1	2	3	4	5	NA
3.6.2. I am able to find out the status of a task/teamwork, e.g. a task in progress, or completion.	1	2	3	4	5	NA
3.7. File/Content Protection						
3.7.1. The system would give me a warning when I try modifying files or notes on the share workspace while my teammates are working on them.	1	2	3	4	5	NA

3.8. Security						
3.8.1. The system seems secure for storing teams' work/files.	1	2	3	4	5	NA
3.8.2. Users need to logon to modify their artifacts or contact their teammates on the system.	1	2	3	4	5	NA

Please circle a number to rank the importance of each criterion of **Collaborativity** to a system's usability (in general).

	not important at all	**a little important**	**important**	**very important**	**Not Applicable**
User Management	1	2	3	4	NA
Awareness	1	2	3	4	NA
Communication	1	2	3	4	NA
User Control/Moderator & Teacher Control	1	2	3	4	NA
File/Content Sharing/Management	1	2	3	4	NA
Process Tracking/Automated Notification	1	2	3	4	NA
File/Content Protection	1	2	3	4	NA
Security	1	2	3	4	NA

Why do you think the criteria of **Collaborativity** above are important or not important to a system's usability (in general)?

Please write down your comments about **Collaborativity of OJS**: (e.g. what make easy or hard to completing a task/team task? what do you like or dislike? What need to be improved on this system?)

4. **Error Tolerance**	**Strongly Disagree**				**Strongly Agree**	**Not Applicable**
4.1. Error Rate (will be collected during observation)						
4.2. Error Prevention						
4.2.1. The system warns me if I am about to make a potential error.	1	2	3	4	5	NA
4.2.2. The system gives me error alerts that clearly tell me how to correct errors.	1	2	3	4	5	NA
4.2.3. Whenever I make a mistake, I am able to recover it easily and quickly e.g. by using an "undo" or "cancel" or "reverse" button.	1	2	3	4	5	NA

Please circle a number to rank the importance of each criterion of **Error Tolerance** to a system's usability (in general).

	not important at all	**a little important**	**important**	**very important**	**Not Applicable**
Error Rate	1	2	3	4	NA
Error Prevention	1	2	3	4	NA

Why do you think the criteria of **Error Tolerance** above are important or not important to a system's usability (in general)?

Please write your comments about **Error Tolerance of OJS**: (e.g. what make easy or hard to completing a task/team task? what do you like or dislike? What need to be improved on this system?)

5. **Universal Accessibility (Ubiquity)**	**Strongly Disagree**				**Strongly Agree**	**Not Applicable**
5.1. Support different users with different levels of IT expertise 5.1.1. The system supports both novice and expert users, advance features are available to expert users.	1	2	3	4	5	NA

Please circle a number to rank the importance of each criterion of **Universal Accessibility (Ubiquity)** to a system's usability (in general).

	not important at all	**a little important**	**important**	**very important**	**Not Applicable**
Support different users	1	2	3	4	NA

Why do you think the criterion of **Universal Accessibility** above is important or not important to a system's usability (in general)?

Please write your comments about **Universal Accessibility of OJS**: (e.g. what make easy or hard to completing a task/team task? what do you like or dislike? What need to be improved on this system?)

6. **Satisfaction**	**Strongly Disagree**				**Strongly Agree**	**Not Applicable**
6.1. Usefulness/Functionality						
6.1.1. This system has all the functions and capabilities I expect it to have.	1	2	3	4	5	NA
6.1.2. The system is useful to my teamwork.	1	2	3	4	5	NA
6.1.3. The various functions in this system are well integrated.	1	2	3	4	5	NA
6.2. Learnability/Predictability/Recognition/Memorability						
6.2.1. It is easy to learn how to use this system.	1	2	3	4	5	NA
6.2.2. Tasks can be performed in a straight-forward manner.	1	2	3	4	5	NA
6.3. Simplicity						
6.3.1. It is simple to use this system.	1	2	3	4	5	NA
6.4. Help/Documentation						
6.4.1. The information (such as online help, on-screen messages, and other documentation) provided on this system is clear, understandable, and helpful.	1	2	3	4	5	NA
6.4.2. It is easy to access help documents.	1	2	3	4	5	NA
6.4.3. I can easily switch between help and my work.	1	2	3	4	5	NA
6.5. . Aesthetic Design						
6.5.1. The interface of this system is pleasant and attractive.	1	2	3	4	5	NA
6.6. Overall						
6.6.1. The system always is reliable.	1	2	3	4	5	NA
6.6.2. I am satisfied with this system.	1	2	3	4	5	NA

Please circle a number to rank the importance of each criterion of **Satisfaction** to a system's usability (in general).

	not important at all	**a little important**	**important**	**very important**	**Not Applicable**
Usefulness/Functionality	1	2	3	4	NA
Learnability/Predictability/Recognition/Memorability	1	2	3	4	NA
Simplicity	1	2	3	4	NA
Help/Documentation	1	2	3	4	NA
Aesthetic Design	1	2	3	4	NA

Why do you think the criteria of **Satisfaction** above are important or not important to a system's usability (in general)?

Please write your comments about **Satisfaction of OJS**: (e.g. what do you like or dislike? Why are you satisfied or not satisfied with the system usability?)

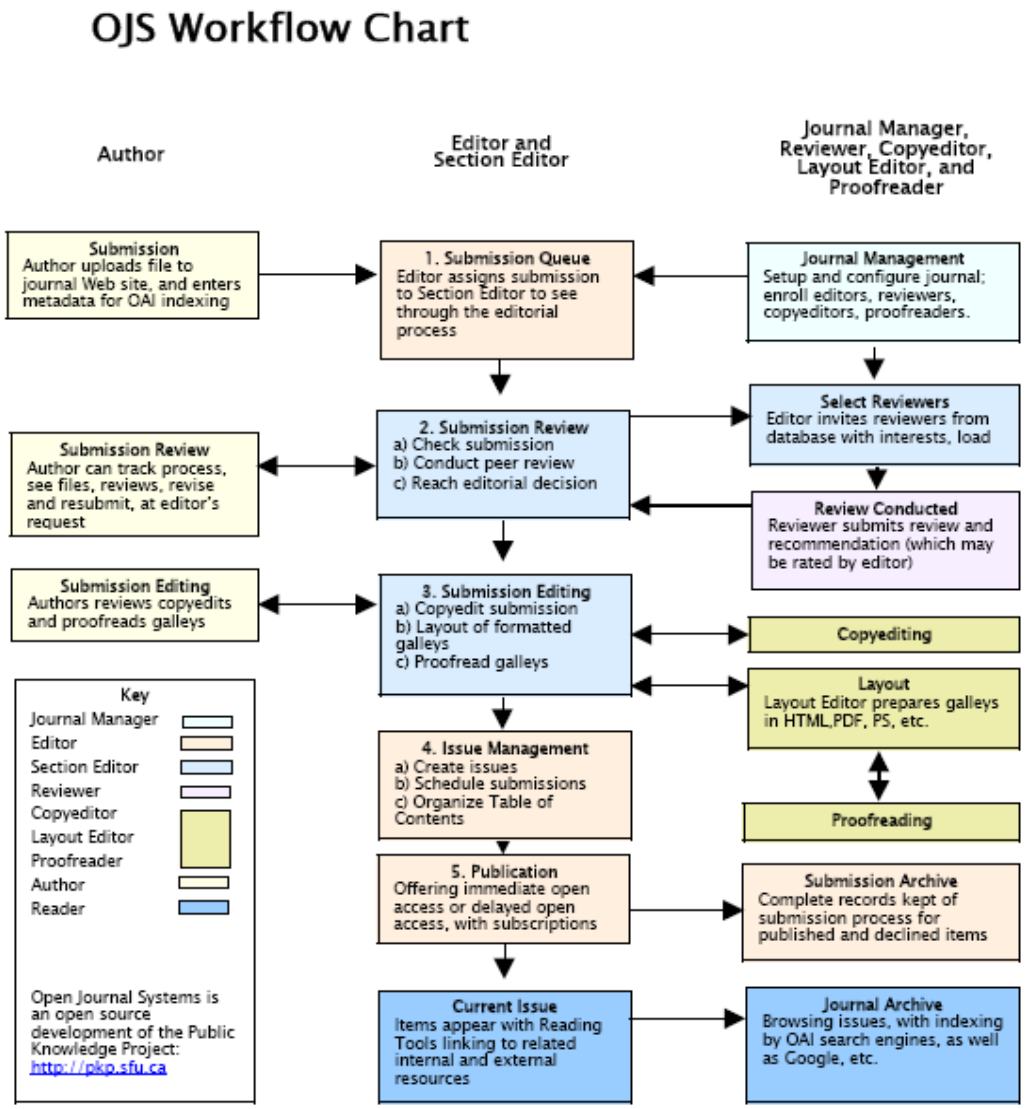
Sources: (Brooke, 1996); (DeBoard, N/A); (González, et al., 2006); (Gutwin & Greenberg, 2000); (Lewis, 1995); (Rubin & Chisnell, 2008); (Shneiderman & Plaisant, 2004); (Xerox-Corporation, 1995);

Approved by the Auckland University of Technology Ethics Committee on 19/May/09.
AUTEC Reference number 09/29

Appendix 5 The list of Interview questions

Q #	Questions
Q1	What do you think of the usability of the system? What do you like or not like, why?
Q2	Regarding to collaborative teamwork, do you think this system is a good system? Why?
Q3	Do you think the system provides enough information, tools/utilities for supporting collaborative teamwork? E.g. decision making. Why do you say so?
Q4	Do you think the criteria shown in the framework are important to UE? What are important or unimportant to the usability of a collaborative learning system?
Q5	Do you think the framework covers mostly important criteria? Why do you say so? What else do you think needed to be included in the framework?
Q6	Suggestions on improvement of OJS, a collaborative teaching and learning system usability in terms of "effectiveness, efficiency, collaborativity, error tolerance, universal accessibility (ubiquity), and satisfaction".
Q7	Communication via email Only, do you think it is ok for collaborative team work?
Q8	Would you like to use this system if you have group assignment/work?
Q9	Overall are you satisfied with OJS?
Q10	What comments/suggestion on the testing do you have?

Appendix 6 OJS Workflow Chart



(PKP, 2008a)

Appendix 7 Collaborative tasks (i.e. multiple users / teamwork) in OJS

Categories (top level)	Roles (multiple users)	Sub-processes (second level)	UE Methods	Note
Journal management	Journal Manager Editor	1) Create new Announcement 2) Create a new section 3) Create Review Form" 4) Select Languages 5) Edit Prepared Email 6) Reading Tools and related Item Tools 7) Statistics and Reports 8) Payments 9) Subscriptions 10) System Plugins 11) Import/Export Data	Questionnaires / Interview (questions & criteria) Screen capture	Work on forms
Review	Editors, or section Editors,	1) Assign a task and send a request to a Reviewer 2) Schedule tasks 3) Track the progress and read the Reviewers comments 4) make a decision – moving on to next stage	Field observation Screen capture Interview / & Questionnaires Heuristic evaluation	change the due date (click the due date link), and send a reminder to the reviewer (use the Send Reminder link),
	reviewers	1) accept the job 2) review the article 3) send a request to reader for revising 4) download the article modified by author, check it again 5) make a decision on accept the submission & send a confirmations to editor/section editor		
	Author	1) submit an article 2) modify the article 3) upload it and send it with an email to Reviewers		
copy editing	Editors, section Editors,	1) Assign a task and send a request to a Copy editor 2) Schedule tasks 3) Track the progress and read the Copy editors' comments 4) make a decision – moving on to next stage	Field observation Screen capture Interview / & Questionnaires Heuristic evaluation	change the due date (click the due date link), and send a reminder to the reviewer (use the Send Reminder link),
	Copy editors	1) Accept the task 2) Check the article and make comments		

		3) Send the file with comments to the author 4) download the article modified by author, check it again 5) Make decision, accept then send an email to Editor /Section Editors		
	Author	1) Download the file and modify it 2) Upload it and send it with an email to Copy editors		
Layout Editing	Editors, section Editors,	1) Assign a task and send a request to a Layout editor 2) Schedule tasks 3) Track the progress and read the Layout editors' comments 4) make a decision – moving on to next stage	Field observation Screen capture Interview / & Questionnaires Heuristic evaluation	change the due date (click the due date link), and send a reminder to the reviewer (use the Send Reminder link),
	Layout editors	1) Accept the task 2) Check the article and make comments 3) Send the file with comments to the author 4) download the article modified by author, check it again 5) Make decision, accept then send an email to Editor /Section Editors		
	Author	3) Download the file and modify it 4) Upload it and send it with an email to Layout editors		
Proofreading	Editors, section Editors,	1) Assign a task and send a request to a Proofreaders 2) Schedule tasks 3) Track the progress and read the Proofreading's comments 4) make a decision – moving on to next stage	Field observation Screen capture Interview / & Questionnaires Heuristic evaluation	change the due date (click the due date link), and send a reminder to the reviewer (use the Send Reminder link),
	Proofreaders	1) Accept the task 2) proofread the article and make comments 3) Send the file with comments to the author 4) download the article modified by author, check it again 5) send the file with comments about layout error to the Layout editors 6) discuss with Editors if needed, then make decision, accept then send an email to Editor /Section Editors		
	Layout Editor	Fix errors based on Proofreaders' comments		
	Author	1) Download the file and modify it 2) Upload it and send it with an email to Proofreaders		

(PKP, 2008a)

Appendix 8 Single user's tasks in OJS

Categories (top level)	Roles (single user)	Sub-processes (second level)	UE method	Note
creating a site	Site Administrator	1) Fill in a form 2) Upload site image or templates 3) Select language + install a new language locale 4) Identify authentication resources	Questionnaires / Interview (questions & criteria) Screen capture	This includes entering the name of your site, an introductory statement about your site, a redirect option (leave this blank if you do not need to redirect users), a description of your site, contact information, a minimum password length for registered users, and indexing registration. Select a language on a form
Site administration	Site Administrator	1) Version Checking 2) Expire User Sessions 3) Clear Data Caches 4) Clear Template Cache 5) Merge Users 6) Counter Statistics	Same as above	
Creating a new journal	Site Administrator	Fill in a form	Same as above	Enter journal title, description, path, etc, enable this journal to appear publicly on your site if needed,
Configuring the new Journal	Journal Manager	5 steps by filling in several forms 1) Fill in Journal's details on a form 2) Define policies on a form 3) Submissions 4) Management 5) The Look	Same as above	**Details**: general info (the name, abbreviation, address, print or online ISSNs, and DOI), Principal Contact, Technical Support Contact, Email Identification, Publisher, Sponsoring Organizations, Sources of Support, Search Engine Indexing; **Policies**: Focus and Scope of Journal, Peer Review (policy, guidelines, process, options), Privacy Statement, Editor Decision, Add Item to Appear in "About the Journal", Journal Archiving, Potential Reviewer Database; **Submissions**: Author Guidelines and Submission Preparation Checklist, Copyright Notice, Competing Interests, For Authors to Index Their Work, Register Journal for Indexing, Notification of Author Submission, **Management**: Access and Security Settings, Publication Scheduling, Identification of Journal Content, Announcements, Copyeditors, Layout Editors, Proofreaders, **The Look**: Journal Homepage Header, Journal Homepage Content, Journal Page Header, Journal Page Footer, Navigation Bar, Journal Layout, Information, Lists
Journal management	Journal Manager Editor	1) Create new Announcement 2) Create a new section 3) Create Review Form"	Same as above	**Announcements**: On a form, select a type, fill in the announcement title, provide the short and detailed descriptions for the announcement, and specify the expiry date for the announcement to display.

		4) Select Languages 5) Edit Prepared Email 6) Reading Tools and related Item Tools 7) Statistics and Reports 8) Payments 9) Subscriptions 10) System Plugins 11) Import/Export Data Work on forms		**Section**: Complete the form with the new section's information, and check the appropriate options, Next, choose a user as the Section Editor. **Review form:** a default form will appear which consists of two text boxes for "author and editor" and for "editor" only; or create a new form where manager can pre-define whether the item is obligatory to be completed to the reviewer, and choose an item type from the dropdown menu e.g. single word text box, extended text box, checkboxes, drop-down box, etc. **Emails**: OJS facilitates work flow communication through the use of internal email messages. The templates for the various messages that are automatically generated can be edited and saved in this section. **Reading tools**: activate and configure them on a form **Payments**: define payment options, Fee Payment Methods on forms, and records **Subscription**: define Subscription Types and Subscription Policies (Expiry, Open Access Options, Author Self-Archiving Policy), and Create New Subscriptions
User Management	Journal Manager	1) Edit existing user account information in this Journal' 2) Enrolling Existing Users 3) Creating Users 4) Merge Users	Same as above	
Create an issue	Journal Editor / Journal manager	1) Create an issue – fill in a form and also define publishing date, 2) Provide table of contents	Same as above	Enter the Volume, Issue, and Year information, select an issue identification format (Year only, Volume/Year, Issue/Volume/Year, etc.), set an Open Access date at this point for the issue as a whole , define the way of Notifying users
Assigning roles / jobs	Editor / Section Editors	Assign users to different roles	Same as above	
Publish/archive issues /articles	Journal Editor / Journal manager	On the Scheduling section, use the dropdown menu; select the issue in which this article will appear.		

(PKP, 2008a)

Appendix 9 Some common tasks can be done by various roles in OJS

Tasks	Roles	Sub-processes
Fill in a form	All roles	1) Enter information on a form; 2) Upload /download files or images; 3) send email for notification
registration	All roles	Fill in a user form
Create new users	Journal manager, site administer, editor, section editor	Fill in user details on a form
Assign roles	Journal manager	Fill in a form – assigning a role
Assign a task	editor, section editor	1) Click a link and then select a role from a user list 2) Send emails to the selected users/roles and the authors 3) Enter instructions / comments
Uploading	Author, editor, section editor, reviewer, copy editor, layout editor, proofreader	1) Upload a file / files 2) Enter or modify metadata on a form 3) Enter descriptions or comments or requirements on the form 4) Send an email for notification and record a decision
Submission Review	Author, editor, section editor, reviewer,	1) Reviewer send an email to editor – accept the request, click the "Will do the review" 2) Click a link to the submission, and Download files 3) Review files and then upload file 4) Fill in the review form 5) Then select to recommendation and submit the review to complete the process. you must enter a review or upload a file before selecting a recommendation
Edit / proofread	Copy editor / editor / section editors / Layout editor /Proofreader	1) Download files; 2) Edit or proofread and make comments /recommendations
Track the editorial progress	Author, editor, section editor, reviewer, copy editor, layout editor, proofreader	Click link and find out the status
Logon	All roles	Enter logon name and password
Log out	All roles	Click logout on user home page
Archive or delete files /issues	Journal manager	1) Browse files and index files; 2) Export and import data
Submit an article	Author	5 steps 1) start - select journal section, check checklist, enter comments, 2) Enter the submission's metadata, Add the title and abstract, Complete the indexing 3) Uploading an article 4) Upload supplementary files if needed 5) Confirming the submission. If the journal requires author fees to be paid, Click on the 'Pay Now' link beside a payment. (note: only mandatory payment at this stage is the Submission Fee)
Pay a fee	Author	Click payment form and enter information needed

(PKP, 2008a)

Appendix 10 Task script

Testing Session (20mins) – one task **Editor/Participant**

Task Title: Peer Review a submission (a group task, 2 users)

The goal of this task: to work out a decision after the two or three Reviewers have reviewed a submitted article - ***The Peer Review Process.doc***.
The Peer Review Team has three users /roles - one Editor, two or three Reviewers.

Task description for the Editor (A participant completes this task)

Logon details:

	Role	**User logon name**	**OJS logon password**	**Email address**
Mark Smith	**Editor**	msmith	msojs09	marksmith-ojs@hotmail.com

The First Round Review
You are Mark Smith, the Editor:

1) open Internet Explorer, go to http://elena.aut.ac.nz/uejmger,
2) enter your logon name and password to logon to the OJS site;

Role assignment

3) Click Editor
4) **Submission**: click Unassigned,
   You will see that one unassigned article which has been submitted by an Author, Helen Singh.
5) Click the file title - ***The Peer Review Process***

On **Summary / Editors**

6) **assign yourself** as the **Editor**;
7) Click **Review** on the top**,** and go to next step – **Peer Review**

On **Peer Review**

8) Click **Select Reviewer**, assign **Ann Webb** and **Rob Will** as the reviewers;
9) click **Mail Icon** to send an email to each reviewer, (please enter your AUT email address as the reviewers' email addresses)
10) a standard email will come up, click **Send**

11) Then logout.

==========

Please wait until you get the indication that you can carry on the following steps

Making a decision
You are Mark Smith, the Editor:

1) Go to http://elena.aut.ac.nz/uejmger, log on to the site if you have not logged on;
2) Click Editor
3) Click **In Review**, you can see the article waiting to be reviewed is highlighted;
4) Click the file title - ***The Peer Review Process;***
5) Check the commends and recommendations from **Ann Webb** (Reviewer) and **Rob Will** (Reviewer);
   Ann Webb's recommendation is "***Decline Submission";***
   Rob Will's recommendation is "***Revisions Required".***
6) On **Editor Decision**: Make a decision, select one choice;

A: If your decision is "*Decline Submission"*, then

7) select the choice - "***Decline Submission";***; click **Record Decision**, click **OK**
8) send an email to notify the **Author, Helen Singh**, with your decision and suggestions.
9) Click **Send**
10) then log out.

--- The end of the task -----

==========

B: if you decide to invite the third person to review the article, then

7) Click **Mail Icon** to send the acknowledgment to **Rob** and **Ann** respectively, a standard email will come up*;*
8) Modify the email if you want, Click **Send;**

On **Editor Decision** Section:

9) select the decision – "***Resubmit for Review"***, click **Record Decision**, click OK
10) upload the article - ***The Peer Review Process.doc*** *(this file is located in "**ojs**" folder on desktop)*;
11) select the uploaded file and click **Resubmit**;

Assign the 3rd reviewer to the Round 2 Peer Review

12) **on Peer Review:** you will notice **Round 1** has been changed to **Round 2**;
13) click **Clear Reviewer** to remove **Ann** and **Rob** from the reviewer list;
14) click **Select Reviewer**, assign **Shirley Tupu** to be the third reviewer for the Round 2;
15) click **Mail Icon** to send **Shirley** an email with Rob and Ann's comments; (*copy Rob's comments from "**accept-resubmit1.doc" in "ojs" folder on desktop**, copy the Ann's comments from **"reject-resubmit1.doc" in "ojs" folder on desktop if needed*)
16) click **Send**
17) then log out.

==========

Please wait until you get the indication that you can carry on the next steps

You are Mark Smith, the **Editor**.

18) log on to the site -http://elena.aut.ac.nz/uejmger;
19) click Editor;
20) Click **In Review**, you can see an article waiting to be reviewed is highlighted;
21) Click the file title - ***The Peer Review Process;***
22) check the commends and the recommendation from Shirley Tupu;
23) make a decision: select a choice; click **Record Decision**, click **OK;**
24) then send an email to notify the Author, Helen Singh, with your decision and suggestions;
25) then log out.

---- The end of the task -----

Task Title: Peer Review a submission (a group task, 2 users) - **Reviewers / Observer**

The goal of this task: to work out a decision after the two or three Reviewers have reviewed a submitted article - ***The Peer Review Process.doc***.
The Peer Review Team has three users /roles - one Editor, two or three Reviewers.

Task description for the Reviewers (the Observer will complete this task)

Logon details:

	Role	**User logon name**	**OJS logon password**	**Email address**
Ann Webb (helper)	**Reviewer**	awebb	awojs09	annwebb-ojs@hotmail.com
Rob Will (helper)	**Reviewer**	rwill	rwojs09	robwill-ojs@hotmail.com
Shirley Tupu	**Reviewer**	stupu	stojs09	shirleytupu-ojs@hotmail.com

The First Round Review

you are **Ann Webb,** a **Reviewer,** have received an email from the Editor.

1) open Internet Explorer, go to http://elena.aut.ac.nz/uejmger,
2) log on to the site;
3) click Reviewer;
4) click the Article Title
5) **Review**: Click **Mail Icon** to send the Editor an email to accept the invitation;
6) a standard email will open up, click **Send** *;*
7) download the article (Submission Manuscript) and save it in "***ojs***" folder on desktop;

8) read the article and make comments (***skip this step due to the time constraint***);
9) click **Review Icon**
10) enter your comments for the Editor (*please copy the comments from **"reject-resubmit1.doc"** in **"ojs"** folder on desktop*), click **Save,** and then click **Close**;
11) upload the file with your commends if you want (*please use the same file you saved on step 7*);
12) select a **Recommendation** – "***Decline Submission",*** then click **Submit review to Editor;**
13) Click OK, an email form will come up;
14) Click **Send***;*
15) then logout.

You are Rob Will, **a Reviewer**

1) repeat the steps above, But
2) copy Rob's comments for the Editor from "***accept-resubmit1.doc"*** in ***"ojs"*** folder on desktop
3) select a Recommendation – "***Revisions Required"***, then click **Submit review to Editor;**
4) Click OK, an email form will come up
5) Click **Send***;*
6) then logout.

==========

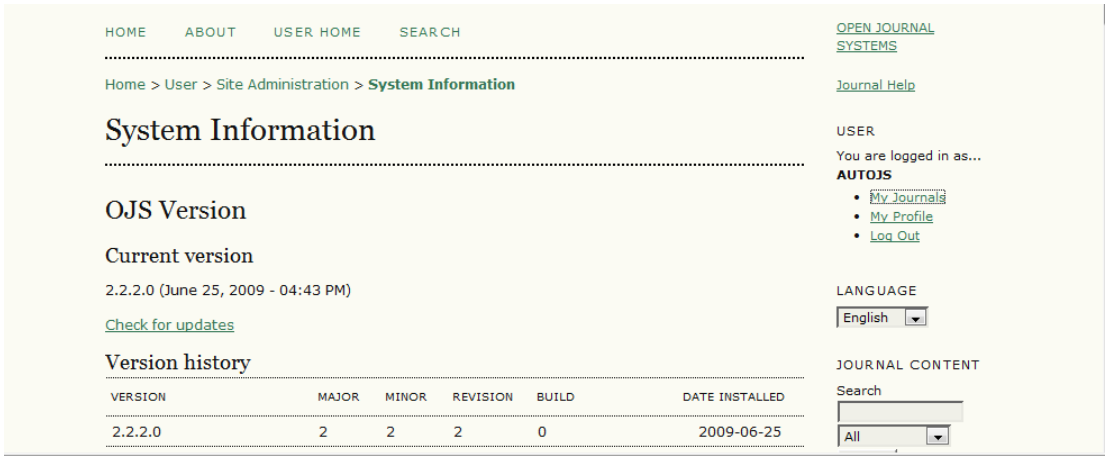
Please wait until you get the indication that you can carry on the second round review

The Second Round Review

you are **Shirley Tupu**, the third **Reviewer**,

1) open Internet Explorer, go to http://elena.aut.ac.nz/uejmger;
2) log on to the site;
3) click Reviewer*;*
4) click the Article Title
5) **Review**: Click **Mail Icon** to send the Editor an email to accept the invitation;
6) a standard email will open up, click **Send***;*
7) download the file and save it in "***ojs***" folder on desktop,
8) read the article and make comments (***skip this step due to the time constraint***),
9) click **Review Icon**
10) enter your comments for the Editor (*please copy the comments from **"accept-resubmit2.doc"** in "**ojs**" folder on desktop*), click **Save** and then click **Close**;;
11) upload the file with your commends if you want (*please use the same file you saved on step 7)*;
7) select a Recommendation – "***Revisions Required,*** then click **Submit review to Editor;**
8) Click OK, an email form will come up
9) Click **Send***;*
10) then logout.

Appendix 11 OJS Journal website for this study (screen-print)



Appendix 12 Approval for AUTEC Ethical Application 09/29

AUT UNIVERSITY
TE WĀNANGA ARONUI O TAMAKI MAKAU RAU

MEMORANDUM

Auckland University of Technology Ethics Committee (AUTEC)

To: Dave Parry
From: **Madeline Banda** Executive Secretary, AUTEC
Date: 19 May 2009
Subject: Ethics Application Number 09/29 **Identifying an effective framework for usability evaluation of Open Journal system in tertiary educational settings.**

Dear Dave
Thank you for providing written evidence as requested. I am pleased to advise that it satisfies the points raised by a subcommittee of the Auckland University of Technology Ethics Committee (AUTEC) at their meeting on 16 March 2009 and that the acting Executive Secretary has approved your ethics application. This delegated approval is made in accordance with section 5.3.2.3 of AUTEC's *Applying for Ethics Approval: Guidelines and Procedures* and is subject to endorsement at AUTEC's meeting on 15 June 2009.
Your ethics application is approved for a period of three years until 19 May 2012.
This approval is for the research instruments as submitted and any alterations need to be submitted for consideration and approved by AUTEC before they are used.
I advise that as part of the ethics approval process, you are required to submit the following to AUTEC:

- A brief annual progress report using form EA2, which is available online through *http://www.aut.ac.nz/about/ethics*. When necessary this form may also be used to request an extension of the approval at least one month prior to its expiry on 19 May 2012;
- A brief report on the status of the project using form EA3, which is available online through *http://www.aut.ac.nz/about/ethics*. This report is to be submitted either when the approval expires on 19 May 2012 or on completion of the project, whichever comes sooner;

It is a condition of approval that AUTEC is notified of any adverse events or if the research does not commence. AUTEC approval needs to be sought for any alteration to the research, including any alteration of or addition to any documents that are provided to participants. You are reminded that, as applicant, you are responsible for ensuring that research undertaken under this approval occurs within the parameters outlined in the approved application.
Please note that AUTEC grants ethical approval only. If you require management approval from an institution or organisation for your research, then you will need to make the arrangements necessary to obtain this. Also, if your research is undertaken within a jurisdiction outside New Zealand, you will need to make the arrangements necessary to meet the legal and ethical requirements that apply within that jurisdiction.
When communicating with us about this application, we ask that you use the application number and study title to enable us to provide you with prompt service. Should you have any further enquiries regarding this matter, you are welcome to contact Charles Grinter, Ethics Coordinator, by email at charles.grinter@aut.ac.nz or by telephone on 921 9999 at extension 8860.
On behalf of the AUTEC and myself, I wish you success with your research and look forward to reading about it in your reports.
Yours sincerely

Madeline Banda
Executive Secretary
Auckland University of Technology Ethics Committee
Cc: Eileen Huang e.huang@auckland.ac.nz, AUTEC Faculty Representative, Design and Creative Technologies

Appendix 13 Invitation to take part in the Usability Evaluation Research Study

Invitation to take part in the Usability Evaluation Research Study

29 June 2009

Dear potential participant,

I would like to invite you to take part in a research study in **usability evaluation of computer supported collaborative learning (CSCL) systems**. I am doing this research as part of my thesis study for completing a master degree in Computer and Information sciences.

The aim of this research is to identify an effective framework for evaluating CSCL system usability. The selected system in this study is **Open Journal System (OJS)** which is a free journal publishing system and has been widely used in many universities around the world. The system is a kind of web-based virtual workplace where scholars and academic staff are able to review and publish e-journals, manage and share knowledge and contents, and collaboratively work together.

As a participant, you are invited to take part in one testing session. Each session will take around 45 – 60 minutes. The following activities in a session will take place:

- You will be observed while working on one or two predefined tasks for testing the system usability
- You will complete one pre-test questionnaire and one post-test questionnaire
- You will be interviewed by me about your satisfaction with the system usability and the identified framework after having finished the task(s) and completed the post-test questionnaire.

A Screen capture application and audio recorders will be used to collect data during a session. Any collected data in this research will remain confidential and anonymous and only used for this thesis and any academic publication that may arise. Any analysis results that would be published will be anonymous.

As with all usability testing, the system is being tested, not you! Please be advised that you will be able to withdraw at any time. There is no advantage in participating and no disadvantage in not participating. The Consent Form will be collected and stored by the SCMS school manager.

If you are willing to participant this research or have questions about this invitation, please reply to me via email (e.huang@auckland.ac.nz) by Friday 17/July/09. Please feel free to distribute my invitation to others if you think they may be interested.

The testing sessions will start from Wednesday 22 July 2009. More details about the test plan will be sent out once I have received your reply.

Your participation and feedbacks about the system usability and the framework will be of great value to my study and I would greatly appreciate if you would accept this invitation.

Yours Sincerely

Eileen Huang
e.huang@auckland.ac.nz

Approved by the Auckland University of Technology Ethics Committee on 19/May/09.
AUTEC Reference number 09/29

Appendix 14 Participant Information Sheet

Participant Information Sheet

Date Information Sheet Produced: 1st May 2009.

Project Title:

Identifying an effective framework for usability evaluation of Open Journal system in tertiary educational settings *(A 90pt thesis)*

An Invitation

My name is Eileen Huang, a Master Student at AUT. I am currently undertaking the research study that will identify an effective framework for usability evaluation (UE) of computer supported collaborative learning (CSCL) systems. I will conduct a system usability testing by applying an identified framework.

You are invited to take part in this research study. Your participation in this study is completely voluntary and you can withdraw from this study at any time without explanation. You also have the right to withdraw any consent given, and to require any data given by you to be destroyed. Whether or not you participate or withdrawal will not in any way affect your grades or academic performances. Your lecturers will not know whether you have taken part in this study.

Before you decide it is important for you to read the information on this sheet and understand why the research is being done and what the process/testing will be. Please take time to read this sheet carefully.

What system will be used for testing?

The system used for usability testing in this study will be Open Journal system (OJS). OJS is a free journal publishing system, and has been widely used in many universities around the world. The system is a kind of virtual workplace where scholars and academic staff are able to publish e-journals, manage and share knowledge and contents, and collaboratively work together.

What is the purpose of this research?

To identify a framework for evaluating the usability of an online system which is used in tertiary collaborative teaching and learning environments.

By using a framework to test a CSCL system (e.g. OJS) usability, I will gather data about user satisfaction with using the system based on the framework criteria and the importance of each framework criterion to the system usability evaluation (UE), then modify the framework, finally conclude an effective framework for a CSCL system UE. **The aim is to measure the usability of a CSCL system, not how good you are at using it !**

The collected data will be used in this study for a thesis in the School of Computing, and Mathematical Sciences, Auckland University of Technology (AUT) and may be in subsequent publications.

How was I chosen for this invitation?

You were chosen because you are

- a postgraduate student, have taken/ will take /are taking the paper – collaborative computing, and currently are not taught by the project Supervisors whose names are shown on page 3;
- or a staff who have been using OJS system at AUT.

What will happen in this research?

You will be a user/participant after you sign the Consent Form. You will attend one testing session. One session will be about 45-60 minute long. The flowchart on page 3 shows the steps and proximate time allotment in a session. During a session, you will be asked to complete a pre-test questionnaire about your past experience in using CSCL system and one usability testing task, fill out a task sheet for the task, then complete a post-test questionnaire regarding your satisfaction with using the system. Finally you will be interviewed about your views on the framework and the system usability.

You will be required to use "**Think Aloud**" protocol while you are doing tasks, i.e. you need to speak out loudly to describe what and why you are going to do and whatever is going through your mind. I will observe you, run screen capture software on the computer which you are working on, and turn on digital recorders while you are doing the task. **Please note that this process tests the system, not you,** and there is no such thing as a wrong answer. So **please try to do whatever you would normally do and be honest in your feedback** – I need to know exactly what you think, not what you think I want to hear.

All you actions and reasons/comments will be recorded and I will take notes while I am observing. Please be advised that you have the option to review, edit or erase the recording. Please ask questions if you need to and let me know when you have finished each task. Then we will take a few minutes to review my notes and your task sheet and questionnaires at the end of the testing session.

What are the discomforts and risk?

There are no known risks in participating. But different people may have different views, understanding, and concerns.

How will these discomforts and risks be alleviated?

You can ask questions if you have some concerns. I will explain and help you understand the questions. Otherwise you could choose to refuse to answer some questions or answer them very briefly.

What are the benefits?

There is no advantage in participating and not disadvantage in not participating.

How will my privacy be protected?

The lecturers who are teaching you and the project supervisors will not be aware of who has participated. The Consent Form will be collected and stored by the SCMS school manager (an independent person).

Any data collected in this research study will remain confidential and anonymous and only used for this thesis and any academic publication that may arise. Any analysis results that would be published will be anonymous. No identifiable information will be made to public, or given to lecturers and the supervisors, etc. You will have the opportunity to exclude information that you do not want to have published.

Once the thesis has been completed, the collected Data from observation and questionnaires will be transferred to a CD/DVD and removed from the computers. The CD/DVD will be stored in a locked filing cabinet in the school post graduate office at AUT Tower building, or other secure site. After six years all the hard copies of data collecting sheets, questionnaires, and the Consent Form will be destroyed.

What are the costs of participating in this research?

About 45-60 minutes of your time for a testing session.

What opportunity do I have to consider this invitation?

As a researcher, I understand that you need time to read this sheet and consider and then get back. I would greatly appreciate it if **you could sign the Consent Form, write down your email address on the form, and bring it into the class next week. I will come and collect it in the classroom WT407.**

How do I agree to participate in this research?

You have indicated that you would like to take part in this research study after you received the Invitation Letter from me. Therefore I give you this **Participant Information Sheet** and the **Consent Form**. If you decide to take part in this study, **please sign the Consent Form, write down your email address on the form, and bring it to the class in WT407 next week**. I will inform you your participant number, when and where you need to come by email once I have received the signed Consent form.

Will I receive feedback on the result of this research?

If you wish to receive a copy of the summary sheet of the testing findings, please indicate it and leave your email address on the Consent Form. I will send it to you once the summary sheet has been completed.

What do I do if I have concerns about this research?

The study has been approved by the Ethics Committee AUT (AUTEC) on 19/May/2009, AUTEC's Ethics Application number is 09/29. Any concerns regarding your rights as a participant and the conduct of the research should be notified to the Executive Secretary, AUTEC, Madeline Banda, email at *madeline.banda@aut.ac.nz* or by phone on 921 9999 extension 8044.

Any concerns regarding the nature of this research should be notified in the first instance to the supervisors.

Whom do I contact for further information about this research?

Researcher

Eileen Huang
Ph: 623 8914
Email: e.huang@auckland.ac.nz

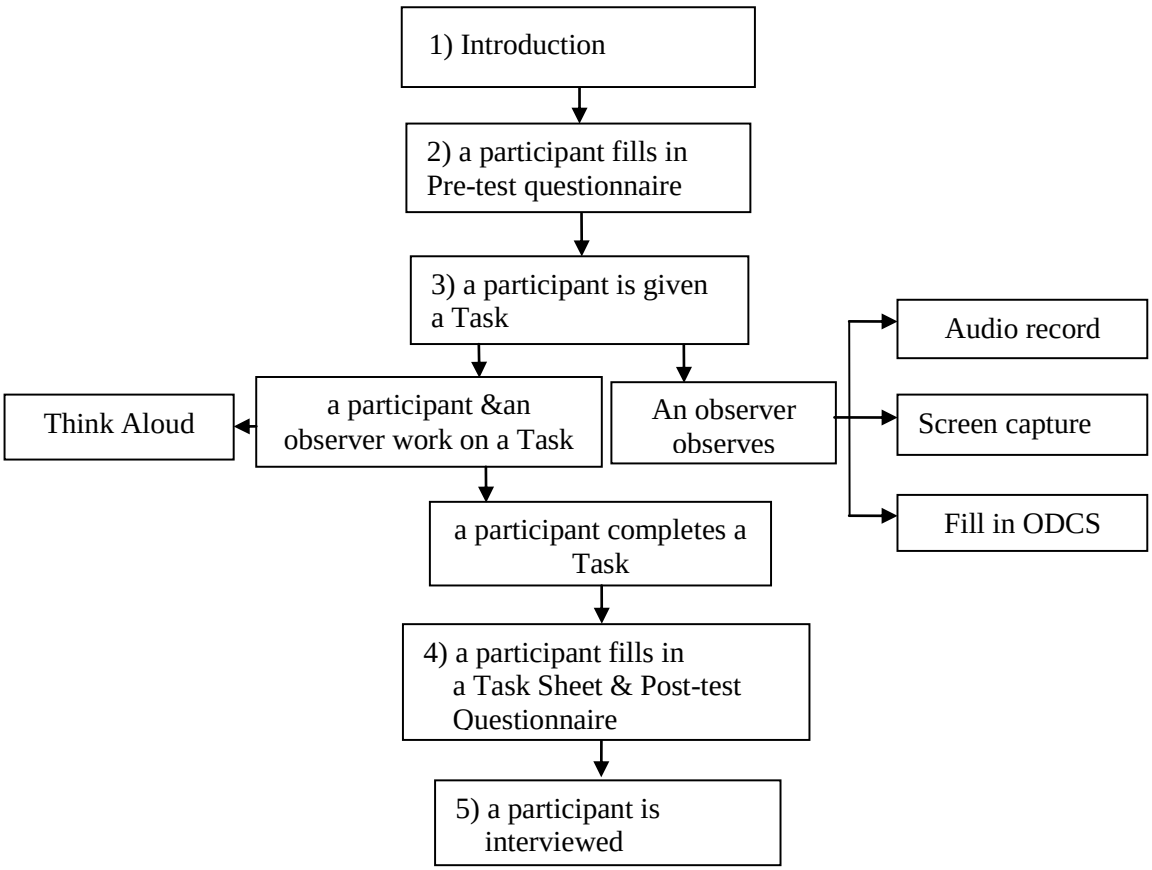
Supervisors

Dr. Dave Parry
Ph: 921 9999 ext 8919
Fax: 9219944
Email: dave.parry@aut.ac.nz

Mali Senapathi
Ph:921 9999 Ext 5213
Fax: 921 9944
Email: mali.senapathi@aut.ac.nz

The Usability Testing Process Flowchart during a session

(up to 60 minutes**)**



Approved by the Auckland University of Technology Ethics Committee on 19/May/09. AUTEC Reference number 09/29

This document is yours to keep.

Appendix 15 Participant Consent Form

Consent Form

Project title: Identifying an effective framework for usability evaluation of Open Journal system in tertiary educational settings

Project Supervisors: Dave Parry (Primary Supervisor) and Mali Senapathi (Second Supervisor)

Researcher: Eileen Huang

Please read carefully and then tick each statement and then sign your name on this form if you agree to participate.

□ I have read and understood the information provided about this research on the Participant Information Sheet dated on 10 May 09.

□ I have had an opportunity to ask questions and to have them answered.

□ I understand that my participation will be recorded on digital audio recorders and captured by Camtasia (a screen capture software) on the computer on which I am working.

□ I understand that data and information I share today will be handled confidentially and anonymously.

□ I understand that my information will be rolled up with the rest of the data from the other study participants. The data may be used for this study and subsequent publications.

□ I understand that I may withdraw myself or any information that I have provided for this project at any time prior to completion of data collection, without being disadvantaged in any way.

□ If I withdraw, I understand that all relevant information including digital audio and video files and transcripts, or parts thereof, will be destroyed.

□ I agree to take part in this research study.

I wish to receive a copy of the summary sheet of the testing findings from the research (please tick one):

Yes□ No□

Participant's signature:_____

Participant's name: _____ Date: _____

Participant's email address (if you would like to have a copy of findings):

……

Approved by the Auckland University of Technology Ethics Committee on 19/May/09.
AUTEC Reference number 09/29
Note: The Participant should retain a copy of this form.

Appendix 16 User Task Sheet (TS)

Usability Testing - User Task Sheet

Introduction on a testing session
You will have one task in this testing Session. While you are doing the task, you are required to **talk loudly** about your actions and reasons for your actions and whatever is going through your mind. All your actions and comments/reasons will be recorded by a digital audio recorder and captured by Camtasia (screen capture software). I will observe and take notes while you are working on the task. Please ask questions if you need to and also indicate when you have finished the task. After you complete the task, please fill out this task sheet and the questionnaires. Then I will take up to 5 minutes to review your notes and my notes with you.

Confidentiality Disclaimer: The collected raw information will be de-identified data and may be used in the thesis and subsequent publications. You have the option to review, edit or erase the records.

Participant #: **Task Title: Peer Review a submission**
Task # Description: *decision making.*

Date: **Starting time:** **completing time:**

1. Have you completed the task successfully? ☐ Yes ☐ No

2. Is this task hard or easy? ☐ Hard ☐ Easy

3. Does the system provide easy ways or tools to help you complete the task? ☐ Yes ☐ No
   The ways / tools are:

4. Did you make mistake when doing the task? ☐ Yes ☐ No
   The mistakes are:

5. Does this task involve team work? ☐ Yes ☐ No
   If yes,
   a) How did you communicate with your teammate via the system? (e.g. email)

   b) Does the system help you collaboratively work with your teammates? ☐ Yes ☐ No
   Why do you say so:

Approved by the Auckland University of Technology Ethics Committee on 19/May/09. AUTEC Reference number 09/29

Appendix 17 Observer Data Collection Sheet (ODCS)

Observer Data Collection Sheet

Confidentiality Disclaimer: The collected raw information will be de-identified data and may be used in the thesis and subsequent publications.
Observer's Name: Eileen Huang **Date**:

Participant number #: **Task Title: Peer Review a submission**

	: Part 1:	**Part 2**	**Part 3:**	**Interview:**
Starting Tim e:				
Completing time:				

1. Has the task been completed? ☐ Yes ☐ No
2. Does the task involve team work? ☐ Yes ☐ No
3. If yes, how many users are there in the team?

4. Number of mistakes: _____
The mistakes are:
Mistakes time spent on correcting it:

5. What questions did the user ask when he/she was working on the task?

6. Problems / issues:

7. Users' Comments on the system usability in term of "effectiveness, efficiency, collaborativity, error prevention, universal accessibility (ubiquity), and satisfaction".

Appendix 18 Time spent in the testing sessions

User#	Part 1 (assign reviewers)	Part 2 (round 1 review)	Part 3: (round 2 review) The final decision	Whole Task (user spent)	Interview
P1	3:43 – 3:56 pm 13min	4:03- 4:39pm 36min	- decided to get the third reviewer but did not completed part 3 (round2)	about 60 min (about 55min) in the middle of round 2	5:10-5:19 pm 9mins
P2	3:39 - 3:45 pm 6 min	3:53 - 3:56 pm 3 min	- declined the submission (round1)	17min (9min) end in round1	4:30 – 4:39pm 9mins
S1	2:25 – 2:29pm 4min	2:38 – 2:47pm 9min	- Declined the submission (round 1)	22min (14min) end in round1	3:20 – 3:30pm 10 min
S2	4:40 - 4:44pm 4min	4:49 – 5:02 pm 13 min	5:04 – 5:07 pm 3 min Revision required (Round2)	27min (20mins) end in round2	5:44- 6:03pm 19 min
S3	2:17 – 2:23pm 5min	2:23 – 2:37 pm 14min	2:38 – 2:43 Pm 5min Revision required (Round2)	27 min (24min) end in round2	3:20 – 3:44pm 24min
S4	2:24 – 2:30 pm 6 min	2:34 – 2:50 pm 16 min	2:52 – 2:57 pm 5 min Revision required (Round2)	33 min (27min) end in round2	3:40 – 3:58pm 18 min
S5	5:49 – 5:51 pm 2 min	5:54 – 6:00pm 6min	6:01 – 6:05 pm 4 min Revision required (Round2)	16 min (12min) end in round2	6:22 – 6:41pm 19min
S6	2:15 – 2:21 pm 6min	2:24 – 2:33 pm 9min	2:34 – 2:38 pm 4 min Accepted the submission (Round2)	23 min (19min) end in round2	2:59 – 3:15pm 16 min
S7	2:03 – 2:10 pm 7 min	2:13 – 2:27 pm 14min	- Accepted the submission (Round1)	24 min (21min) end in round1	3:59 – 4:11pm 12 min
S8	2:04 – 2:11 pm 7min	2:15 – 2:18 pm 3 min	- Revision required (Round1)	14 min (10min) end in round1	2:40 – 3:00pm 20 min
S9	5:37 – 5:42 pm 6min	5:45 – 5:49 pm 4min	- Declined the submission (Round1)	12 min (10min) end in round1	6:30 - 6:45pm 15 min
S10	2:07 – 2:11 pm 4min	2:14 – 2:23 pm 9 min	2:24 – 2:29 pm 5 min Revision required (Round2)	22 min (18min) end in round2	3:25 – 3:42pm 17min

Note: P standards for the participants in pilot study and S standards for the participants who were students from the CC class.

Appendix 19 TG & FU's Internal Reliability (IR) Statistics – Part 1

Note for the tables:

- number of cases: is the number of the participants, "valid" means the responses from the participants did not include N/A (Not applicable) or N/R (No response).
- number of Items : means the number of statements.

TG & FU's IR Statistics – Part 1

	Measures	Cronbach's Alpha		Number of Items		Valid Number of Cases	
	Groups	TG	FU	TG	FU	TG	FU
Dimensions	**Effectiveness**	.855	.783	6	6	11	10
	Efficiency	.859	.889	8	8	7	9
	Collaborativity	.658	.409	12	10	3	3
	Error Tolerance	.766	.938	3	3	6	9
	Universal Accessibility	-	-	-	-	-	-
	Satisfaction	.476	.955	10	12	3	7

Note: Each of the following component variables has zero variance and is removed from the scale: Q3.1.1. User Management, Q3.6.1. Process Tracking/Automated Notification, Q3.7.1. File/Content Protection.
Each of the following component variables has zero variance and is removed from the scale: Q6.2.2. Learnability/Predictability/Recognition/Memorability, Q6.4.2. Help/Documentation.
The determinant of the covariance matrix is zero or approximately zero. Statistics based on its inverse matrix cannot be computed and they are displayed as system missing values.

TG & FU's IR Statistics – Part 1

	Groups	TG		FU	
	Measures	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Statement# / Criterion	Q1.1.1. /Completeness	.339	.902	.237	.820
	Q1.2.1. /Visibility	.864	.784	.502	.765
	Q1.3.1. /Organisation/Design	.427	.865	.678	.723
	Q1.3.2. /Organisation/Design	.819	.796	.491	.762
	Q1.4.1./ Navigability	.836	.800	.576	.739
	Q1.4.2. /Navigability	.734	.820	.830	.673
	Q2.1.1. /Speed	.797	.824	.699	.876
	Q2.1.2. /Speed	-.121	.902	.169	.909
	Q2.2.1. /Familiarity/ Consistency/Standards	.625	.841	.656	.878
	Q2.2.2. /Familiarity/ Consistency/Standards	.404	.874	.895	.848
	Q2.3.1./ Flexibility/ Adaptability/Configurability	.573	.846	.685	.873
	Q2.4.1. /Effort	.881	.803	.838	.855
	Q2.4.2. /Effort	.959	.799	.520	.893
	Q2.4.3. /Effort	.818	.815	.867	.853
	Q3.1.1. /User Management	.866	.516	-	-
	Q3.1.2. /User Management	.991	.460	.000	.492
	Q3.2.1. /Awareness	.500	.619	-.115	.474
	Q3.3.1. /Communication	.305	.640	-.115	.474
	Q3.4.1. /User Control/ Moderator & Teacher control	-.500	.720	-.115	.474
	Q3.4.2. /User Control /Moderator & Teacher control	-	-	.866	.094

	Q3.5.1. /File/Content Sharing & Management	.500	.619	.693	.173
	Q3.5.2. /File/Content Sharing & Management	.500	.619	.693	.173
	Q3.6.1./ Process Tracking /Automated Notification	-.292	.738	-	-
	Q3.6.2. /Process Tracking /Automated Notification	-.292	.738	.693	.173
	Q3.7.1. /File/Content Protection	.189	.663	-	-
	Q3.8.1. /Security	.397	.618	-.115	.474
	Q3.8.2. /Security	.866	.516	-.327	.536
	Q4.2.1. /Error Prevention	.731	.529	.948	.856
	Q4.2.2. /Error Prevention	.882	.430	.881	.907
	Q4.2.3. /Error Prevention	.316	.960	.800	.965
	Q6.1.1. /Usefulness/ Functionality	.500	.352	.470	.959
	Q6.1.2. /Usefulness/ Functionality	.971	.173	.463	.959
	Q6.1.3. /Usefulness/ Functionality	.971	.173	.962	.946
	Q6.2.1. /Learnability/Predictability/ Recognition/Memorability	.115	.474	.873	.949
	Q6.2.2. /Learnability/Predictability/ Recognition/Memorability	-	-	.901	.948
	Q6.3.1. /Simplicity	-.500	.630	.898	.949
	Q6.4.1. /Help/Documentation.	-.500	.630	.963	.946
	Q6.4.2. /Help/Documentation.	-	-	.712	.954
	Q6.4.3. /Help/Documentation	-.756	.683	.712	.954
	Q6.5.1. /Aesthetic Design	.500	.281	.847	.950
	Q6.6.1 /Overall Satisfaction	.971	.173	.860	.949
	Q6.6.2. /Overall Satisfaction	.500	.352	.830	.952

Appendix 20 TG & FU's Internal Reliability (IR) Statistics – Part 2

Note for the tables:

- number of cases: is the number of the participants, "valid" means the responses from the participants did not include N/A (Not applicable) or N/R (No response).
- number of Items : means the number of criteria.

TG & FU's IR Statistics – Part 2

	Measures	Cronbach's Alpha		Number of Items		Valid Number of Cases	
	Groups	TG	FU	TG	FU	TG	FU
Dimensions	Effectiveness	.207	-.865	4	4	12	13
	Efficiency	-.012	.303	4	4	12	13
	Collaborativity	.803	.628	8	8	10	11
	Error Tolerance	.667	.807	2	2	11	10
	Universal Accessibility	-	-	-	-	-	-
	Satisfaction	.755	.748	5	5	10	13

a. The value is negative due to a negative average covariance among items. This violates reliability model assumptions. You may want to check item codings.
The determinant of the covariance matrix is zero or approximately zero. Statistics based on its inverse matrix cannot be computed and they are displayed as system missing values.TG & FU's Reliability

TG & FU's IR Statistics – Part 2

	Groups	TG		FU	
	Measures	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Question# / Criterion	Q1a1 /Completeness	-.133	.495	-.371	-.238[a]
	Q1a2 /Visibility	.051	.222	-.585	.176
	Q1a3 /Organisation/Design	.362	-.308[a]	-.081	-1.714[a]
	Q1a4 /Navigability	.219	.020	.099	-1.969[a]
	Q2a1 /Speed	-.028	.039	-.031	.472
	Q2a2 /Familiarity/Consistency/ Standards	.193	-.412[a]	.175	.227
	Q2a3 /Flexibility/Adaptability/ Configurability	-.137	.156	.252	.132
	Q2a4 /Effort	-.040	.051	.336	.117
	Q3a1 /User Management	.579	.771	.023	.657
	Q3a2 /Awareness	.628	.770	.636	.521
	Q3a3 /Communication	.000	.846	.183	.631
	Q3a4 /User Control/Moderator & Teacher Control	.591	.770	.040	.688
	Q3a5 /File/Content Sharing/Management	.410	.797	.572	.516
	Q3a6 /Process Tracking /Automated Notification	.591	.770	.696	.472
	Q3a7 /File/Content Protection	.844	.721	.772	.438
	Q3a8 /Security	.633	.769	-.183	.710
	Q4a1 /Error Rate	.516	.[a]	.701	.[a]
	Q4a2 /Error Prevention	.516	.[a]	.701	.[a]
	Q6a1 /Usefulness/Functionality	.510	.717	.350	.754

	Q6a2 /Learnability/Predictability/ Recognition/Memorability	.796	.615	.456	.724
	Q6a3 /Simplicity	.816	.605	.664	.641
	Q6a4 /Help/Documentation	.124	.877	.672	.642
	Q6a5 /Aesthetic Design	.580	.689	.448	.726
	a. The value is negative due to a negative average covariance among items. This violates reliability model assumptions. You may want to check item coding.				

Appendix 21 Inter-Rater Reliability (IRR) Analysis Statistics – Part 1

Note for the tables:
Table -"Case Processing Summary":
"Case N" means the number of cases, i.e. number of statements.
Table - "Reliability Statistics":
"N of Items" means the number of participants.

IRR Analysis Statistics – Part 1

Case Processing Summary - Cases N

	Groups	TG			FU		
	Cases N	**Valid**	**Excluded[a]**	**Total**	**Valid**	**Excluded[a]**	**Total**
Dimensions	Effectiveness	4	2	6	4	2	6
	Efficiency	6	2	8	7	1	8
	Collaborativity	4	9	13	5	8	13
	Error Tolerance	-	-	-	-	-	-
	Universal Accessibility	-	-	-	-	-	-
	Satisfaction	7	5	12	6	6	12

Note: *a. Listwise deletion based on all variables in the procedure.*

Reliability Statistics

	Groups	TG		FU	
	Measures	Cronbach's Alpha	N of Items	Cronbach's Alpha	N of Items
Dimensions	Effectiveness	.536	12	.604	13
	Efficiency	.861	12	.491	13
	Collaborativity	.882	12	.848	13
	Error Tolerance	-	-	-	-
	Universal Accessibility	-	-	-	-
	Satisfaction	.055	12	-.086	13

Intraclass Correlation Coefficient (Average Measures)

	Groups	TG	FU
	Measures	Intraclass Correlation[a]	Intraclass Correlation[a]
Dimensions	Effectiveness	.536[b]	.604[b]
	Efficiency	.861[b]	.491[b]
	Collaborativity	.882[b]	.848[b]
	Error Tolerance	-	-
	Universal Accessibility	-	-
	Satisfaction	.055[b]	-.086[b]

Note: *Two-way mixed effects model where people effects are random and measures effects are fixed.*
a. Type C intraclass correlation coefficients using a consistency definition-the between-measure variance is excluded from the denominator variance.
b. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

Appendix 22 Inter-Rater Reliability (IRR) Analysis Statistics – Part 2

Note for the tables:
Table -"Case Processing Summary":

- "Case N" means the number of cases, i.e. number of criteria.

Table - "Reliability Statistics":

- "N of Items" means the number of participants.

IRR Analysis Statistics– Q2

Case Processing Summary - Cases N

	Groups	TG			FU		
	Cases N	**Valid**	**Excluded[a]**	**Total**	**Valid**	**Excluded[a]**	**Total**
Dimensions	Effectiveness	4	0	4	4	0	4
	Efficiency	4	0	4	4	0	4
	Collaborativity	6	2	8	5	3	8
	Error Tolerance	-	-	-	-	-	-
	Universal Accessibility	-	-	-	-	-	-
	Satisfaction	4	1	5	5	0	5

Note: *a. Listwise deletion based on all variables in the procedure.*

Reliability Statistics

	Groups	TG		FU	
	Measures	Cronbach's Alpha	N of Items	Cronbach's Alpha	N of Items
Dimensions	Effectiveness	-1.493[a]	12	.542	13
	Efficiency	-.545 [a]	12	.564	13
	Collaborativity	.533	12	.319	13
	Error Tolerance	-	-	-	-
	Universal Accessibility	-	-	-	-
	Satisfaction	.854	12	.227	13

Note: *a. The value is negative due to a negative average covariance among items. This violates reliability model assumptions. You may want to check item codings.*

Intraclass Correlation Coefficient (Average Measures)

	Groups	TG	FU
	Measures	Intraclass Correlation[a]	Intraclass Correlation[a]
Dimensions	Effectiveness	-1.493[b]	.542[b]
	Efficiency	-.545[b]	.564[b]
	Collaborativity	.533[b]	.319[b]
	Error Tolerance	-	-
	Universal Accessibility	-	-
	Satisfaction	.854[b]	.227[b]

Note: *Two-way mixed effects model where people effects are random and measures effects are fixed.*

a. Type C intraclass correlation coefficients using a consistency definition-the between-measure variance is excluded from the denominator variance.

b. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

Appendix 23 Frequency Percentage Comparison between TG and FU – Statement rankings – Part 1

Level of Agree-ment		Not Applicable (N/A)		No Response (N/R)		Strongly Disagree 1		Disagree 2		Undecided 3		Agree 4		Strongly Agree 5	
%/#		%/#		%/#		%/#		%/#		%/#		%/#		%/#	
	Groups	TG	FU	TG	FU	TG%	FU	TG	FU	TG	FU	TG	FU	TG	FU
Statements on OJS	Q1.1.1	0/0	0/0	0/0	0/0	17/2	0/0	0/0	8/1	33/4	23/3	42/5	46/6	8/1	23/3
	Q1.2.1	0/0	0/0	0/0	0/0	0/0	8/1	25/3	15/2	25/3	31/4	33/4	46/6	17/2	0//0
	Q1.3.1	0/0	0/0	0/0	0/0	8/1	0/0	17/2	23/3	67/8	46/6	8/1	31/4	0/0	0/0
	Q1.3.2	0/0	0/0	8/1	8/1	8/1	0/0	17/2	8/1	25/3	31/4	42/5	54/7	0/0	0/0
	Q1.4.1	0/0	0/0	0/0	0/0	0/0	8/1	42/5	8/1	33/4	46/6	25/3	31/4	0/0	8/1
	Q1.4.2	0/0	0/0	8/1	23/3	8/1	0/0	42/5	31/4	33/4	23/3	8/1	23/3	0/0	0/0
	Q2.1.1	0/0	0/0	0/0	0/0	0/0	0/0	17/2	0/0	33/4	31/4	50/6	46/6	0/0	23/3
	Q2.1.2	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	8/1	23/3	50/6	62/8	42/5	15/2
	Q2.2.1	0/0	0/0	0/0	0/0	0/0	0/0	25/3	8/1	58/7	46/6	17/2	38/5	0/0	8/1
	Q2.2.2	0/0	0/0	0/0	0/0	0/0	0/0	8/1	15/2	41/5	23/3	25/3	31/4	25/3	31/4
	Q2.3.1	42/5	31/4	0/0	0/0	0/0	0/0	0/0	8/1	25/3	15/2	25/3	31/4	8/1	15/2
	Q2.4.1	8/1	0/0	0/0	0/0	0/0	15/2	17/2	0/0	25/3	38/5	25/3	38/5	25/3	8/1
	Q2.4.2	0/0	0/0	0/0	0/0	0/0	8/1	0/0	8/1	42/5	46/6	33/4	31/4	25/3	8/1
	Q2.4.3	0/0	0/0	0/0	0/0	0/0	0/0	0/0	8/1	50/6	31/4	17/2	38/5	33/4	23/3
	Q3.1.1	17/2	62/8	0/0	0/0	0/0	0/0	8/1	0/0	8/1	31/4	42/5	8/1	25/3	0/0
	Q3.1.2	8/1	54/7	0/0	8/1	0/0	0/0	0/0	0/0	17/2	23/3	33/4	8/1	42/5	8/1
	Q3.2.1	0/0	0/0	0/0	0/0	8/1	8/1	33/4	23/3	25/3	23/3	25/3	38/5	8/1	8/1
	Q3.3.1	8/1	8/1	0/0	0/0	0/0	0/0	0/0	15/2	50/6	38/5	33/4	38/5	8/1	0/0
	Q3.4.1	17/2	38/5	0/0	0/0	8/1	0/0	0/0	0/0	33/4	23/3	42/5	38/5	0/0	0/0
	Q3.4.2	33/4	8/1	8/1	15/2	0/0	0/0	0/0	0/0	8/1	23/3	42/5	46/6	8/1	8/1
	Q3.5.1	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	17/2	15/2	50/6	46/6	33/4	38/5
	Q3.5.2	25/3	0/0	0/0	0/0	0/0	0/0	0/0	8/1	17/2	23/3	33/4	31/4	25/3	38/5
	Q3.6.1	0/0	0/0	0/0	0/0	0/0	0/0	0/0	8/1	8/1	8/1	58/7	38/5	33/4	46/6
	Q3.6.2	0/0	0/0	0/0	0/0	0/0	0/0	8/1	0/0	17/2	0/0	67/8	62/8	8/1	38/5
	Q3.7.1	67/8	46/6	0/0	0/0	8/1	0/0	0/0	8/1	17/2	38/5	0/0	8/1	8/1	0/0
	Q3.8.1	50/6	8/1	0/0	0/0	0/0	8/1	17/2	0/0	17/2	46/6	8/1	15/2	8/1	23/3
	Q3.8.2	33/4	15/2	0/0	8/1	0/0	0/0	8/1	0/0	8/1	15/2	50/6	31/4	0/0	31/4
	Q4.2.1	25/3	15/2	0/0	0/0	33/4	8/1	8/1	23/3	8/1	38/5	25/3	8/1	0/0	8/1
	Q4.2.2	42/5	15/2	0/0	0/0	33/4	8/1	8/1	23/3	17/2	23/3	0/0	23/3	0/0	8/1
	Q4.2.3	33/4	23/3	0/0	0/0	25/3	8/1	8/1	15/2	25/3	31/4	8/1	8/1	0/0	15/2
	Q5.1.1	8/1	8/1	8/1	8/1	0/0	0/0	25/3	38/5	17/2	23/3	42/5	23/3	0/0	0/0
	Q6.1.1	0/0	0/0	0/0	0/0	0/0	8/1	0/0	0/0	25/3	38/5	67/8	46/6	8/1	8/1
	Q6.1.2	0/0	0/0	0/0	0/0	0/0	0/0	8/1	0/0	25/3	38/5	58/7	54/7	8/1	8/1
	Q6.1.3	17/2	0/0	0/0	0/0	0/0	0/0	0/0	15/2	33/4	46/6	42/5	31/4	8/1	8/1
	Q6.2.1	0/0	0/0	0/0	0/0	0/0	8/1	17/2	23/3	25/3	8/1	42/5	31/4	17/2	31/4
	Q6.2.2	0/0	0/0	0/0	8/1	0/0	0/0	8/1	23/3	17/2	23/3	67/8	31/4	8/1	15/2
	Q6.3.1	0/0	0/0	0/0	8/1	0/0	8/1	25/3	8/1	25/3	15/2	33/4	31/4	17/2	31/4
	Q6.4.1	42/5	0/0	0/0	0/0	0/0	0/0	8/1	23/3	33/4	46/6	17/2	23/3	0/0	8/1
	Q6.4.2	75/9	15/2	0/0	0/0	0/0	0/0	0/0	0/0	25/3	46/6	0/0	31/4	0/0	8/1
	Q6.4.3	75/9	23/3	0/0	0/0	0/0	0/0	0/0	0/0	17/2	46/6	8/1	23/3	0/0	8/1
	Q6.5.1	0/0	0/0	0/0	8/1	0/0	8/1	33/4	15/2	33/4	31/4	17/2	23/3	17/2	15/2
	Q6.6.1	33/4	8/1	0/0	0/0	0/0	8/1	0/0	8/1	50/6	31/4	17/2	38/5	0/0	8/1
	Q6.6.2	0/0	0/0	0/0	0/0	0/0	0/0	17/2	15/2	25/3	31/4	58/7	46/6	0/0	8/1

Appendix 24 Std Dev, Max, Min, & Mean Comparison between TG and FU – Statement rankings (Part 1)

		Std. Dev		Max		Min		Mean	
	Groups	TG	FU	TG	FU	TG	FU	TG	FU
Statements on OJS	Q1.1.1	1.22	0.90	5	5	1	4	3.25	3.85
	Q1.2.1	1.08	0.99	5	4	2	3	3.42	3.15
	Q1.3.1	0.75	0.76	4	4	1	3	2.75	3.08
	Q1.3.2	1.04	0.67	4	4	1	4	3.09	3.50
	Q1.4.1	0.83	1.01	4	5	2	3	2.83	3.23
	Q1.4.2	0.82	0.88	4	4	1	3	2.45	2.90
	Q2.1.1	0.78	0.76	4	5	2	4	3.33	3.92
	Q2.1.2	0.65	0.64	5	5	3	4	4.33	3.92
	Q2.2.1	0.68	0.78	4	5	2	3	2.88	3.46
	Q2.2.2	1.03	1.09	5	5	2	4	3.63	3.77
	Q2.3.1	0.76	0.97	5	5	3	4	3.71	3.78
	Q2.4.1	1.12	1.17	5	5	2	3	3.64	3.23
	Q2.4.2	0.83	1.01	5	5	3	3	3.83	3.23
	Q2.4.3	0.94	0.93	5	5	3	4	3.83	3.77
	Q3.1.1	0.94	0.45	5	4	2	3	4.00	3.20
	Q3.1.2	0.79	0.89	5	5	3	4	4.27	3.60
	Q3.2.1	1.16	1.14	5	5	1	3	2.92	3.15
	Q3.3.1	0.69	0.75	5	4	3	3	3.55	3.25
	Q3.4.1	0.95	0.52	4	4	1	4	3.30	3.63
	Q3.4.2	0.58	0.63	5	5	3	4	4.00	3.80
	Q3.5.1	0.72	0.73	5	5	3	4	4.17	4.23
	Q3.5.2	0.78	1.00	5	5	3	4	4.11	4.00
	Q3.6.1	0.62	0.93	5	5	3	4	4.25	4.23
	Q3.6.2	0.75	0.51	5	5	2	4	3.75	4.38
	Q3.7.1	1.63	0.58	5	4	1	3	3.00	3.00
	Q3.8.1	1.17	1.17	5	5	2	4	3.17	3.50
	Q3.8.2	0.74	0.79	4	5	2	4	3.63	4.20
	Q4.2.1	1.41	1.08	4	5	1	3	2.33	2.82
	Q4.2.2	0.95	1.18	3	5	1	3	1.71	3.00
	Q4.2.3	1.16	1.29	4	5	1	3	2.25	3.10
	Q5.1.1	0.92	0.87	4	4	2	3	3.20	2.82
	Q6.1.1	0.58	0.97	5	5	3	3	3.83	3.46
	Q6.1.2	0.78	0.63	5	5	2	4	3.67	3.69
	Q6.1.3	0.67	0.85	5	5	3	3	3.70	3.31
	Q6.2.1	1.00	1.39	5	5	2	4	3.58	3.54
	Q6.2.2	0.75	1.08	5	5	2	3	3.75	3.42
	Q6.3.1	1.08	1.29	5	5	2	4	3.42	3.75
	Q6.4.1	0.69	0.90	4	5	2	3	3.14	3.15
	Q6.4.2	0.00	0.69	3	5	3	4	3.00	3.55
	Q6.4.3	0.58	0.71	4	5	3	4	3.33	3.50
	Q6.5.1	1.11	1.22	5	5	2	3	3.17	3.25
	Q6.6.1	0.46	1.07	4	5	3	3	3.25	3.33
	Q6.6.2	0.79	0.88	4	5	2	3	3.42	3.46

Appendix 25 Frequency Percentage Comparison between TG and FU– Importance rankings (Part 2)

Level of importance		Not Applicable (N/A)		No Response (N/R)		Not important 1		A little important 2		Important 3		Very important 4	
%/#		%/#		%/#		%/#		%/#		%/#		%/#	
	Groups	TG	FU	TG	FU	TG	FU	TG	FU	TG	FU	TG	
Criteria	Q1a1	0/0	0/0	0/0	0/0	0/0	0/0	8/1	0/0	25/3	46/6	67/8	54/7
	Q1a2	0/0	0/0	0/0	0/0	0/0	0/0	0/0	8/1	58/7	77/10	42/5	15/2
	Q1a3	0/0	0/0	0/0	0/0	0/0	0/0	8/1	8/1	50/6	38/5	42/5	54/7
	Q1a4	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	58/7	38/5	42/5	62/8
	Q2a1	0/0	0/0	0/0	0/0	0/0	0/0	17/2	8/1	42/5	38/5	42/5	54/7
	Q2a2	0/0	0/0	0/0	0/0	0/0	0/0	8/1	15/2	25/3	31/4	67/8	54/7
	Q2a3	0/0	0/0	0/0	0/0	0/0	0/0	0/0	15/2	50/6	69/9	50/6	15/2
	Q2a4	0/0	0/0	0/0	0/0	0/0	0/0	8/1	8/1	50/6	85/11	42/5	8/1
	Q3a1	0/0	8/1	0/0	0/0	0/0	0/0	0/0	0/0	42/5	54/7	58/7	38/5
	Q3a2	0/0	8/1	0/0	0/0	0/0	0/0	8/1	8/1	67/8	62/8	25/3	23/3
	Q3a3	0/0	0/0	0/0	0/0	0/0	0/0	0/0	8/1	42/5	15/2	58/7	77/10
	Q3a4	0/0	0/0	0/0	0/0	0/0	0/0	0/0	23/3	33/4	38/5	67/8	38/5
	Q3a5	0/0	0/0	0/0	0/0	0/0	0/0	0/0	15/2	25/3	38/5	75/9	46/6
	Q3a6	0/0	0/0	0/0	8/1	0/0	0/0	0/0	15/2	42/5	31/4	58/7	46/6
	Q3a7	0/0	0/0	8/1	0/0	0/0	0/0	8/1	15/2	50/6	23/3	33/4	62/8
	Q3a8	0/0	0/0	17/2	0/0	0/0	0/0	17/2	8/1	8/1	15/2	58/7	77/10
	Q4a1	0/0	15/2	8/1	8/1	0/0	0/0	0/0	8/1	42/5	38/5	50/6	31/4
	Q4a2	0/0	15/2	0/0	0/0	0/0	0/0	0/0	0/0	17/2	31/4	83/10	54/7
	Q5a1	0/0	0/0	0/0	23/3	8/1	0/0	0/0	23/3	50/6	15/2	42/5	38/5
	Q6a1	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	33/4	62/8	67/8	38/5
	Q6a2	0/0	0/0	0/0	0/0	0/0	0/0	0/0	8/1	42/5	54/7	58/7	38/5
	Q6a3	0/0	0/0	0/0	0/0	0/0	0/0	0/0	23/3	58/7	38/5	42/5	38/5
	Q6a4	0/0	0/0	17/2	0/0	0/0	0/0	8/1	15/2	33/4	62/8	42/5	23/3
	Q6a5	0/0	0/0	0/0	0/0	0/0	0/0	8/1	15/2	75/9	69/9	17/2	15/2

Appendix 26 Std Dev, Max, Min, & Mean Comparison between TG and FU – Importance rankings (Part 2)

Measures		Std. Dev		Max		Min		Mean	
Groups		TG	FU	TG	FU	TG	FU	TG	FU
Criteria	Q1a1	0.67	0.52	4	4	2	3	3.58	3.54
	Q1a2	0.51	0.49	4	4	3	2	3.42	3.08
	Q1a3	0.65	0.66	4	4	2	2	3.33	3.46
	Q1a4	0.51	0.51	4	4	3	3	3.42	3.62
	Q2a1	0.75	0.66	4	4	2	2	3.25	3.46
	Q2a2	0.67	0.77	4	4	2	2	3.58	3.38
	Q2a3	0.52	0.58	4	4	3	2	3.50	3.00
	Q2a4	0.65	0.41	4	4	2	2	3.33	3.00
	Q3a1	0.51	0.51	4	4	3	3	3.58	3.42
	Q3a2	0.58	0.58	4	4	2	2	3.17	3.17
	Q3a3	0.51	0.63	4	4	3	2	3.58	3.69
	Q3a4	0.49	0.80	4	4	3	2	3.67	3.15
	Q3a5	0.45	0.75	4	4	3	2	3.75	3.31
	Q3a6	0.51	0.78	4	4	3	2	3.58	3.33
	Q3a7	0.65	0.78	4	4	2	2	3.27	3.46
	Q3a8	0.85	0.63	4	4	2	2	3.50	3.69
	Q4a1	0.52	0.67	4	4	3	2	3.55	3.30
	Q4a2	0.39	0.50	4	4	3	3	3.83	3.64
	Q5a1	0.87	0.92	4	4	1	2	3.25	3.20
	Q6a1	0.49	0.51	4	4	3	3	3.67	3.38
	Q6a2	0.51	0.63	4	4	3	2	3.58	3.31
	Q6a3	0.51	0.80	4	4	3	2	3.42	3.15
	Q6a4	0.70	0.64	4	4	2	2	3.40	3.08
	Q6a5	0.51	0.58	4	4	2	2	3.08	3.00

Appendix 27 The problems occurred when students used OJS to work on group assignments in Collaborative Computing class