

ADAPTING COLLABORATIVE LEARNING TOOLS TO SUPPORT GROUP PEER MENTORSHIP

A Thesis Submitted to the
College of Graduate Studies and Research
in Partial Fulfillment of the Requirements
for the degree of Doctor of Philosophy
in the Department of Computer Science
University of Saskatchewan
Saskatoon

By

Oluwabunmi (Adewoyin) Olakanmi

© Oluwabunmi (Adewoyin) Olakanmi October 2016 All rights reserved

PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Head of the Department of Computer Science
176 Thorvaldson Building
110 Science Place
University of Saskatchewan
Saskatoon, Saskatchewan
Canada
S7N 5C9

ABSTRACT

Group peer mentorship is a relatively new addition to the area of collaborative learning. We see an untapped potential in supporting this model of mentorship with the existing collaborative learning tools like peer review and wiki. Therefore, we proposed to use a modified peer review system and a modified wiki system. From our preliminary studies using both peer review and wiki systems, we found that participants preferred the peer-review system to the wiki system in supporting them for mentorship. Therefore, this dissertation specifically addresses how to adapt the peer review system to support group peer mentorship.

We proposed a modified peer review system, which comprises seven stages – initial submission of the first draft of the paper by the author, the review of author’s paper by peer reviewers, release of review feedback to the author, back-evaluation of their reviews by the authors, modification of the paper by the author, submission of the final paper and the final stage where both authors and reviewers provide an evaluation of the peer review process with respect to their learning, their perception of the helpfulness of the process, and their satisfaction with the process. We also proposed to use our group matching algorithm, based on some constraints and the principles of the Hungarian algorithm, to achieve a diversified grouping of peers for each peer review session. With these, we conducted six peer review studies with the graduate and undergraduate students at the University of Saskatchewan and teachers in Chile. This dissertation reports on the findings from these studies.

We found that peer review, with some modifications, is a good tool to facilitate group peer mentorship. An evaluation of the performance of our group matching algorithm showed an improvement over three other algorithms, with respect to three metrics – knowledge gain of peers, time and space consumption of the algorithm. Finally, this dissertation also shows that wiki has the potential to support group peer mentorship, but needs further research.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my supervisor, Dr Julita Vassileva, for her support and encouragement during my doctoral research. Thanks for your patience and willingness to help every step of the way. I am incredibly grateful to you for the funding and all the opportunities I had, as your student. I couldn't have been successful in my PhD journey without your support. Thanks for being a great mentor and advisor.

I would like to appreciate my PhD committee members – Dr Gord McCalla, Dr Jim Greer and Dr Keith Walker. I truly appreciate your time, feedback, suggestions and the guidance that I received in the course of my doctoral study. I have learned a lot from your questions and insightful comments. My gratitude also goes to Dr Amali Weerasinghe for bringing her expertise in the area of collaborative learning to judge my thesis. Special thanks to Gwen Lancaster for her help and support throughout my PhD journey.

My appreciation goes to Dr Fred Phillips, Dr Chris Dutchyn and Dr Chanchal Roy for allowing me to run my studies in their classes. Also, thanks to the team at the Educational Research Center of the University of Chile – Roberto Araya, Paulina Sepúlveda Guzmán and Josefina Hernandez for the opportunity to run my studies with the teachers in Chile. I also appreciate Francisco Gutierrez for his help with translating our results from Spanish to English. Also, thanks to Dr Kewen Wu for his help with the Wiki study. Special thanks to Dr Akindele Odeshi for his support on this journey. I really appreciate all the words of encouragement and the re-assuring that all will be well, even in the face of difficult situations.

Thanks to my fellow graduate students in MADMUC lab and other departments at the University of Saskatchewan, Johnson Iyilade, Richard Lomortey, Rita Orji, Shaikhah Alotaibi, Edgar Leilei, Bukola Ishola, Ifeoma Adaji, Kiemute Oyibo, Marwah Heidi, Adeola Olubamiji, Ayodele Olagunju and Aniekan Udofia; for their stimulating discussions and the fun we had in the last few years of my studies.

Special appreciation goes to my parents, late Chief and Mrs S. A. Adewoyin, for the opportunity they afforded me in life. Thanks for all the sacrifices you made. Also, thanks to my siblings for their support.

Finally, I will like to specially appreciate my husband, Moses Olakanmi, who has been continually supportive of my doctoral study. Thank you for your sacrifices, unconditional love and encouragement through what I can describe as the most challenging years of my life. Special thanks to my daughter, Oluwasemilore, for enduring this journey with me. Thanks for being my sunshine on the seemingly dark days.

TABLE OF CONTENTS

Permission To Use	i
Abstract	ii
Acknowledgments	iii
Table Of Contents	v
List Of Tables	viii
List Of Figures	x
1 Introduction	1
1.1 Problem.....	2
1.1.1 How do mentees learn?	3
1.1.2 What models of e-learning can support online mentoring?	4
1.1.3 Optimal group formation and mentor-to-mentee group recommendation	5
1.1.4 How to motivate participation.....	6
1.2 Approach.....	7
1.3 Contributions.....	8
1.4 Dissertation outline	10
2 Related Work	11
2.1 Group Peer Mentorship.....	11
2.1.2 Classification of Mentorship Systems.....	13
2.1.3 Peer Mentoring	16
2.1.4 Factors that can affect mentoring relationship	18
2.2 Computer Supported Collaborative Learning.....	19
2.2.2 Theories of Collaboration.....	20
2.2.3 Technology Support for Collaborative Learning	25
2.3 Peer review in group learning.....	26
2.3.1 Learning domains where peer review can be used	27
2.3.2 Criticisms of peer review	29
2.3.3 Existing work to improve on the peer review process	30
2.4 Wiki in collaborative learning	32
2.4.1 Classification of wiki editors.....	33
2.4.2 How to motivate wiki participation.....	33
2.5 General Issues in Collaborative Learning.....	35
2.6 Group matching in collaborative learning	37
2.7 Summary.....	45
3 Thesis Statement	46
3.1 Research Questions.....	46
3.2 Answering the Research Questions	47

3.3	Proposed Approach.....	49
3.3.1	Modified Peer Review Process	49
3.3.2	Modified Wiki.....	51
3.3.3	Group formation and review assignment.....	52
4	Adapting Online Peer Review To Support Group Peer Mentorship	55
4.1	Introduction.....	55
4.2	Methodology.....	57
4.2.1	Implementing group peer mentorship	57
4.2.2	Feedback Loop	58
4.3	Case Studies, Results and Discussion.....	59
4.3.1	Exploratory Study - Ethics and IT, 2013.....	60
4.3.2	Case Study - Ethics and IT, 2014.....	68
4.3.3	Comparison of findings from the two experiments (Ethics 2013 and Ethics 2014) .	71
4.3.4	Case Study - Research Methods (CMPT 880/890) Class, 2013-2014	72
4.3.5	Case Study – Ethics & IT Class, 2016	75
4.3.6	Case Study – Masters of Professional Accounting Class, 2015.....	79
4.4	Conclusion	83
5	Large Scale Case Study With Teachers In Chile	86
5.1	Introduction.....	86
5.2	Research Method	87
5.3	Case Study Results and Discussion – Mathematics Group	89
5.4	Case study results and discussion for combined results	94
5.5	Conclusion	96
6	Matching Peers In Group Peer Mentorship System.....	98
6.1	Problem Definition.....	98
6.1.1	Group formation algorithms in collaborative learning.....	99
6.1.2	Limitations of the existing group formation tools.....	102
6.2	Group formation as akin to assignment problem in Mathematics	103
6.2.1	The Hungarian Algorithm	103
6.2.2	Matching in the Stable Marriage Problem	106
6.2.3	Tournament Organization	107
6.3	Proposed Matching Algorithm.....	108
6.4	Sample peer matching from our algorithm	110
6.5	Manual Time and Space Analyses of our Algorithm.....	112
6.6	Experimental Test of the Peer Matching Algorithm.....	112
6.6.1	Datasets	113
6.6.2	Evaluating the performance of the peer matching algorithm.....	113
6.6.3	Aggregate Knowledge Gains by Peers	114
6.6.4	Time and Space Analysis	115

6.7	Limitations of our algorithm.....	117
6.8	Conclusion	117
7	Alternative Tool To Support Group Peer Mentorship – Use Of Wiki.....	119
7.1	Introduction.....	119
7.2	Research Tool	119
7.3	Experiment.....	122
7.4	Results and Discussion	123
7.5	Conclusion	128
8	Contributions And Conclusion.....	130
8.1	Discussion of Our Findings	130
8.1.1	Research Question 1	130
8.1.2	Research Question 2.....	131
8.1.3	Research Question 3.....	132
8.1.4	Research Question 4.....	133
8.2	Contributions.....	134
8.3	Implications of our Findings on Collaborative Learning.....	136
8.4	Limitations and Future Work.....	138
8.5	Conclusion	139
	My Peer-Reviewed Publications With Contents From This Dissertation	140
	References	142
	Appendix A	168
	Part 1: Total gain from the groups	168
	Part 2: Total time consumption of each algorithm.....	169
	Part 3: Total space consumption of each algorithm.....	170
	Appendix B	171
	Exit Questionnaire (CMPT 408 – Winter 2013 & 2014 / CMPT 880 – Fall 2013/Winter 2014)	
	171
	Appendix C	182
	Part 1: Informed Consent Form – MPAcc – Fall 2015 / CMPT 408 – Winter 2016.....	182
	Part 2: Rotter’s Interpersonal Trust – Distrust Scale	184
	Part 3: Back-evaluation Survey (MPAcc – Fall 2015 / CMPT 408 – Winter 2016)	190
	Part 4: Peer Review Session Survey (MPAcc – Fall 2015 / CMPT 408 – Winter 2016).....	194
	Appendix D	197
	Behavioural Ethics Approval Certificates	197

LIST OF TABLES

Table 2.1: Learning domains where peer review systems can be used	28
Table 2.2: Existing group formation tools in collaborative learning	39
Table 4.1: Overview of the 5 case studies	59
Table 4.2: Authors' ratings of the reviews they received. The ratings are expressed as level of agreement with the statements expressing the criteria, using a scale of 1 (lowest) to 10 (highest))	62
Table 4.3: Marks given to the peers' final essay by the marker	64
Table 4.4: Marks given to the peers' reviews by the marker	65
Table 4.5: Satisfaction levels of authors with the reviews they received according to the Exit Questionnaire (students' agreement with statements expressing the criteria is reported on a scale of 1 (lowest) to 10 (highest))	69
Table 4.6: Peers' grades from their final essay	70
Table 4.7: Satisfaction levels of authors with the reviews they received (agreement with statements expressing the criteria, using a scale of 1 (lowest) to 10 (highest)).....	73
Table 4.8: Progression of participants' grades as judged by the marker.....	76
Table 4.9: Overall evaluation of the weekly essays by the peer reviewers	77
Table 4.10: Average helpfulness of the weekly peer reviews	78
Table 4.11: Correlation between weekly review ratings and back-evaluation ratings - Ethics and IT, 2016.....	79
Table 4.12: Correlation: trust scores vs. back-evaluation & trust scores vs. average review ratings	82
Table 4.13: Mean and Standard deviation of trust scores and ratings for each quartile of trust scores.....	82
Table 4.14: Correlation: trust scores vs. individual review ratings	83
Table 5.1: Average peer ratings of the participants' papers from the first peer review	90
Table 5.2: Review feedback from the second peer review session.....	90
Table 5.3: Paired-Samples T-test on the Data from the second peer review session	91

Table 5.4: Back-evaluation of the reviews from the first peer review session	91
Table 5.5: Back-evaluation of the reviews from the second peer review session	92
Table 5.6: Paired-Samples T-test on the Data from the Second Peer Review Session	93
Table 5.7: Helpfulness and Learning from the Peer Review Experiment	93
Table 5.8: Average lengths of the reviews & standard deviation; Correlation btw back-evaluation & length of review; R-Squared (R^2)	95
Table 5.9: Correlation between back-evaluation and review ratings (merging reviewers)	96
Table 7.1: Grades assigned to individual contributions.....	124
Table 7.2: Weekly contribution quality for each student.....	125
Table 7.3: Participants' perception of other contributors	126
Table 7.4: Participants' preference for motivation strategies.....	127

LIST OF FIGURES

Figure 3.1: Framework of the modified peer review system	51
Figure 4.1: Authors' satisfaction with the reviews received in each reviewing session: (a)-on a weekly basis, (b)-in the final questionnaire.	64
Figure 4.2: Authors' satisfaction with the reviews received in each reviewing session (Research methods, 2014).....	74
Figure 4.3: Participants' interpersonal trust score from Rotter's scale	81
Figure 5.1: Distribution of participants from the Mathematics teachers' group	88
Figure 6.1: Total gain in peers` abilities from the three datasets (Four algorithms compared)..	115
Figure 6.2: Time to run the four algorithms compared.....	116
Figure 6.3: Space usage by the four algorithms compared	116
Figure 7.1: Sample email notification.....	121
Figure 7.2: Dialogue box showing content changes & acceptance/evaluation of changes	121
Figure 7.3: Grades given to the final articles for the four weeks.....	124

CHAPTER 1

INTRODUCTION

With the advent of ICT tools, learning has gone beyond the four walls of classrooms. In the last decade, most educational institutions have adopted the use of e-learning tools to either support or replace their existing learning system. Some of the benefits they derive from adopting e-learning technologies are flexible and collaborative learning opportunities for learners, which are enhanced through synchronous and asynchronous communication.

Mentorship is a relationship between a mentor, who is expected to be more knowledgeable and experienced, and a mentee, who is being guided by the mentor for either career or psychological development (Ragins and Kram, 2007). In traditional mentorship, mentors are expected to be older and more experienced. However, research had shown that peer mentorship is a strong source of peers' cognitive and affective developments (Astin, 1996). Peer mentorship is a reciprocal peer support and learning situation whereby peer mentors help their peers to achieve their academic, career or psychosocial goals, and they all eventually acquire transferrable skills. In this mode of mentorship, the peer mentor is believed to be close in age, career or academic achievement to the mentee, and can help to enhance and promote the overall experience of the mentees as friend and learning facilitator (Pompper and Adams, 2006; Clark *et al.*, 2011). Group peer mentorship is a form of collaborative learning, which brings together in a group peers with similar goals, for either long or short term period, and is dissolved once their goals are achieved or in case of other reasons for dissolution (e.g. lack of group support for effective mentoring). Also, group peer mentorship is not always just a one-time learning activity, where peers are matched and disengaged once the learning activity is concluded. Here, peers' goals can be wide and far-reaching, ranging from learning, career development to rendering psychological support, and can range from short to long term mentoring relationship. Research had described factors that can affect mentoring relationship as communication, commitment, training and support for participants, understanding the peculiarity of participants' situations and trust (Hodges, 2009; Thomas, 2011; Martins and Sifers, 2012) (see section 2.1.4).

Different platforms supporting learning at large scale and collaboration exist at the moment, for example MOOCs (massive open online courses). However, research had shown that

the completion rate in MOOCs is low, due to lack of motivation on the part of learners, learners' intention to take course for credit or for audit purpose, lack of individual feedback on concepts and techniques in the course, economic and technological factors (DeVine, 2013; Ho *et al.*, 2014; Reich, 2014; Pretz, 2014). Some of the solutions proposed to enhance success in such learning platforms are providing strong motivation for learners to participate in and finish the course, enhancing learners' communication and collaboration with their instructors and their peers, and allowing learners to take central role in evaluation (DeVine, 2013; Jordan, 2013).

Particularly, communication is important because in this way the students can receive feedback by humans, but not necessarily by automated systems. Personal feedback is expensive and impractical in MOOCs since they can have tens of thousands of students. So a possible solution is to engage students in providing feedback to each other, by mentoring each other, answering questions, and evaluating each other's' work. An argument for this is given by Anstin (1996). An argument against peer-evaluation specifically has been given by Jordan (2013). Yet, the cost and the benefits outweigh the potential disadvantages. We believe that through a peer feedback system and feedback from experienced mentors, learners can get direction. Therefore, we propose a group peer mentorship system, where peers are mentors and mentees to each other and a senior peer provides a quality control of the feedback provided by peers.

1.1 Problem

Group Peer mentorship, according to Minor (2007), is an ideal form of collaborative venture where peers are both mentors and mentees. This research focuses on online group peer mentorship system. Many online mentorship platforms exist at the moment e.g. MentorNet¹, CyberMentor², BlueCare E-mentoring system³. However, we found that existing mentorship systems typically do manual pairing of mentors and mentees in one-to-one mentorship system. Implementing this system requires a one-to-one matching of mentors and mentees, which is

¹ <http://www.mentornet.net/>

² <http://www.cybermentors.ca/>

³ Stewart, S. and McLoughlin, C. (2007). Design features of an e-mentoring system for the health professions: Choosing to learn in partnership. The Australasian Society for Computers in Learning in Tertiary Education (ascilite). ICT: Providing Choices for Learners and Learning. ISBN 978-981-05-9579-1 2-5. December 2007, Singapore; pp. 951-961.

expensive for two reasons. First, experienced mentors are hard to come by, and their time is valuable. Second, the matching process is expensive, since a lot of information needs to be known about their goals, preferences, interests, etc. in advance and there is no guarantee that a good match will be found for any mentee. A less costly solution to this problem is group matching of mentors and mentees. Research findings have shown that group mentorship has some advantages over the one-to-one mentorship, which the existing mentorship sites typically do (Herrera et al, 2002; Lawrence et al, 2008). However with the group mentorship, we still need to get dedicated experienced mentors, whose time is costly. Therefore, we propose group peer mentorship, where peers mentor each other in groups. While there are no dedicated experienced mentors, the natural differences in the knowledge and experiences in a group of peer learners facilitate knowledge transfer (Vygotsky, 1978). With few notable exceptions (Greer *et al.*, 1997) group peer mentorship model has not been deployed widely in e-learning systems (Adewoyin and Vassileva, 2013). The following sub-sections discuss the research issues pose by the proposed mentorship model.

1.1.1 How do mentees learn?

People want to be mentored to satisfy their career needs, for some specific course needs or personal goals. For example, some people need to receive guidance on how to navigate their career path while others require guidance to manage other pertinent situations. So, how do mentees learn? First, it is very important for mentors to elicit what the past accomplishments, current state and beliefs of mentees are, when initiating mentoring relationship. Knowing these can help to reinforce mentees' belief in their ability and aid the success of the mentoring relationship. For example, a mentee with a history of career failure does not need a mentor that will make him feel like a total failure, but someone that can link his past success with his choices, in order to identify where his strength lies. The Cognitivists believe that one of the factors that can either enhance or inhibit learning is the individual's pre-existing knowledge of the world (Schunk, 2008). Therefore, the first thing to do in mentoring is to be able to identify the existing states and knowledge levels of mentees, in order to know what intervention strategy will be most suitable for them.

Second, they can learn by observing and imitating other mentees and their mentors; which is also a point in support of group peer mentorship. This is what Robert Sears and Albert

Bandura called “observational learning” (Sears, 1951; Bandura, 1977). For example in the medical profession, mentees (physicians-in-training) learn by observing their consultants while on training (Sears, 1951; Grusec, 1992). According to the socio constructivists, interaction among peers is a powerful source of knowledge transfer (Piaget, 1928). That is, they can learn through interaction with their peers and their ability to take into consideration others’ perspectives (Lipponen, 2002). Also, according to the socio-cultural theorists, social interaction influences peers’ development if it occurs within their zone of proximal development. Zone of proximal development is defined as *"the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance, or in collaboration with more capable peers"* (Vygotsky, 1978). Therefore, in group peer mentorship it is essential to have peers participate in a heterogeneous group, with a mix of capabilities, through which they can complement each other and the natural differences in their knowledge and experiences can facilitate knowledge transfer and enhance learning and cognitive development.

1.1.2 What models of e-learning can support online mentoring?

E-learning enhances flexible and collaborative learning among learners and tutors. It has evolved, over the years, from a model that gives learners flexible access to learning materials online and many other kinds of pedagogical interactions (content, service and technology model) (MacDonald *et al.*, 2001) to an instructional design model that supports collaborative learning (Conrad, 2000). Online mentorship systems are a relatively new addition to e-learning. In mentorship systems, learning is motivated when there is collaboration between mentees and mentors. Lev Vygotsky believed that children learn by interacting with their environment. The environment comprises their peers and their mentors, who are often the same people (Vygotsky, 1978). Jean Piaget also believed that children learn more when they interact with their peers in contrast to an adult giving them instructions (Piaget, 1928; Long *et al.*, 2011). Therefore, mentees will benefit more from group peer mentorship, where they can collaborate with their peers and leverage each other’s strengths, rather than from one-to-one mentorship.

Tools used in collaborative learning include computer applications such as email, bulletin board, peer review systems and wikis. Most of these applications are not designed with learning in mind, but we believe that online group peer mentorship systems can be supported with these

collaborative tools to enhance interaction and creativity among participants, give structure to their tasks, and also provide storage for personal and shared data (Lipponen, 2002; Adewoyin and Vassileva, 2013).

1.1.3 Optimal group formation and mentor-to-mentee group recommendation

Participants in online mentorship sites have different interests and goals. Matching mentors to group of mentees is not straightforward in group mentorship systems. First, how do we ensure that each group models the preferences and features of the constituent members? In Masthoff (2004) and Masthoff (2010), some group modeling strategies were discussed. These are plurality voting, utilitarian strategy, Borda count, Copeland rule, approval voting, least misery strategy, average without misery, most pleasure strategy, most respected person strategy and fairness strategy. These strategies are useful for aggregating preferences of individual group members, which will be used for group formation. Another problem is recommending a group of mentors to group of mentees. In McCarthy and Anagnost (1998), O'Conner *et al.* (2001), Masthoff (2004) and Masthoff and Gatt (2006), some group recommendation algorithms were used to recommend items like movies, video and music clips to group participants.

In this research, we are considering group peer mentorship. Therefore, the problem is not about recommending group of mentors to group of mentees, but to find the perfect combination of peers in a group, in which they can successfully mentor one another. Winter and McCalla (2003), in their experiment with student teams working on software engineering projects, described four factors that could affect group performance – knowledge and functional skills, teamwork behaviors, task type, and contextual and situational factors. In this experiment, the group marks was used as the outcome measure of the group success. They found that the functional task skill of the group members is the most important predictor of the success (marks) of the student project groups. Therefore, it is important to ensure that, within each group, peers have the required knowledge requirements for the group task. In our case, it is important to have peers with varying skills that are needed to help one another for successful mentorship. In addition, the scenario here is that of matching peers in groups such that some form of reciprocity is involved. Greer *et al.* (1998), Pizzato *et al.* (2010) and Yu *et al.* (2011) used the concept of reciprocal recommendation in their research on recommending people to people, but not in groups. Also, researchers in the field of computer supported collaborative learning (CSCL) have

proposed different group matching algorithms (see section 2.6). However, some of these algorithms still rely on manual matching to support their strategies, consider limited constraints in the matching and rely on self-report to evaluate the performance of the algorithms (Vivacqua and Lieberman, 2000; Redmond, 2001; Tobar *et al.*, 2004). Therefore, we believe that a group matching algorithm, which does not rely solely on manual matching, allows for changing grouping criteria and can be evaluated using a combination of the system data and self-report, will be a good solution.

1.1.4 How to motivate participation

A collaborative e-learning system will be a failure with no quality contribution by users because the core of any such online community is the content contributed by users through active participation. Although some research findings have shown that group mentoring gives more opportunities for promoting positive interaction among the participants than one-to-one mentoring (Herrera *et al.*, 2002; Lawrence *et al.*, 2008), it can also give rise to lurking. A lurker can be described as someone who only browses, learns from and gets entertained with what others are doing in the community. However, it is important for every participant in a mentorship system to be actively involved, in order for their peers and mentors to understand what their goals and problems are and how to help them.

In Albert Bandura's social learning theory, self-regulation and self-efficacy encourage learning and development of certain behaviors. Self-regulation occurs without expecting assessment from external sources by using goal setting and self-assessment (Bandura, 1986). Therefore, providing ways by which mentees can set goals and monitor their progress towards achieving their goals will not only help them to achieve their goals but also motivate them to be active in the forum. Self-efficacy has to do with individuals' belief in their ability, which can be enhanced by a history of past personal achievements, and training and feedback from group members (Bandura, 1977b; Bandura, 1986). In addition, behaviorists and social learning theorists generally believe that learners get motivated when they are offered appropriate rewards for their actions (Ormrod, 1999; Allen, 2007). Farmer and Glass (2010), and Hamilton *et al.* (2011) described vote to promote, content rating and ranking, content reviewing and comments, quality karma and incentive points as some of the tools for providing feedback of individual's activities in the group. Therefore, we believe that offering appropriate rewards for participation

will encourage participation, strengthen participants' confidence in their ability and also encourage proper behavior.

1.2 Approach

This research focuses on the problem of how to support effectively group peer mentorship through designing an appropriate mechanism and tool, and evaluating its effectiveness experimentally in case studies. The proposed mechanism in this thesis is modified peer review. The traditional peer review mechanism, used for scientific journal, comprises five stages: initial submission of the first draft of the paper by the author, the review of author's paper by peer reviewers, usually three peer reviewers; release of review feedback to the author, modification of the paper by the author, and submission of the final paper. The modified peer review mechanism includes an additional stage of providing back-evaluation of their reviews by the authors, which happens immediately after their review feedback is released to them, and a final stage where both authors and reviewers provide an evaluation of the peer review process with respect to their learning, their perception of the helpfulness of the process, and their satisfaction with the process. In addition, peers are grouped for each peer review session based on our group matching strategy using some concepts of the traditional Hungarian algorithm. Hungarian algorithm is a combinatorial optimization algorithm which is used to solve assignment problem in Mathematics. We see the problem of group matching as a bipartite graph, which is a special case of k -graph with the vertices decomposed into two disjoint sets such that vertices within the same sets are not adjacent (Weisstein, 2014). Therefore, we see peers as the disjoint sets which can be matched to form mentorship group as in a bipartite graph, such that the grouping does not contradict our matching constraints.

Before we decided to choose this mechanism, we did an extensive literature review of online collaborative learning tools, specifically peer review and wiki systems, collaborative theories, interpersonal trust, and mentorship. We experimented with the use of peer review and wiki systems to support our mentorship model.

1. With the modified peer review system, review assignments were done not randomly but in groups of up to four peers with mixed abilities, so that weaker peers could learn from stronger ones by operating within their zone of proximal development. Double-blind

review was used in each group to avoid bias and reciprocation over different review sessions. In addition, we believe that this would also ensure objective and high-quality reviews. Peers as authors back-evaluated the peer review they got of their work. At the end of each peer review session, peers evaluated the learning they gained and the general helpfulness of the session in improving their skills both as authors and reviewers. To measure the helpfulness of the modified peer review system to the authors, we analysed their back-evaluation ratings and comments as well as their overall feedback on their learning and the system's general helpfulness.

2. With the wiki system, peers worked collaboratively to produce a single document on a given topic. In this system, peers were able to edit the document by adding, deleting, or rearranging the contents. However, peers could also undo and rate any change(s) made to their contributions in the wiki system. In this wiki system, we included some new features like email notification for every change made to a peer's contribution, the highlighting of changes made to their contribution, the opportunity to accept or deny changes made to their contribution, and the computation of peers' reputations based on the acceptance or denial of their contributions.

When considering the best tool for our group peer mentorship, we conducted experiments with both peer review and wiki systems in an undergraduate computer science class, where participants used both systems for their course work. We found, from the log and qualitative data, that the participants preferred the peer review system to the wiki system. Therefore, we followed the modified peer review framework to support the group peer mentorship model in this thesis.

1.3 Contributions

The goal of this research is to find an appropriate mechanism to support group peer mentorship. The contribution of this thesis is in the area of collaborative learning with an immediate impact on learning in educational settings, on professionals engaging in lifelong learning, and on the scientific community enhancing the credibility of the peer review system. These impacts are discussed as follows:

1. Introduction of the back-evaluation stage to the peer review process, by allowing authors to provide feedback on the perceived helpfulness and the quality of the reviews that they receive. We believe that the back-evaluation will serve as an incentive for the reviewers to do a better job.
2. Development of a group-matching algorithm. This dissertation provides an evaluation of the matching algorithm and compares it with three other grouping algorithms used in grouping peers for learning purposes: the random, stratified, and grouping algorithm by Agrawal *et al.* (2014). We found that our algorithm saves execution time and space and enhances peers' skills in collaborative learning when compared with the three group matching algorithms mentioned above.
3. Identification of effective measures of the success of peer review sessions. This dissertation evaluates the success of the peer review session using two metrics; the helpfulness of the peer feedback to the authors, using both log and qualitative data; and peers' satisfaction with the peer review process. We also have a designated marker, who grades the quality of their reviews and their final papers, as a senior peer. We look at their grades for each peer review session as provided by the senior peer. The grades constitute a useful measure of learning for both peer reviews and the collaborative learning environments.
4. Identification of key motivation strategies to enhance the success of the proposed group peer mentorship. This dissertation discusses both intrinsic and extrinsic motivation strategies to encourage peers to give thorough and helpful reviews and unbiased back-evaluation.
5. Clarification of review length as a weak predictor of review quality. This dissertation demonstrates the weak correlation between review length and the review quality, using peers' feedback and the system log data. However, this finding is still somewhat preliminary needing further confirmation.
6. Clarification of the relationship between peer review ratings and the back-evaluation ratings each subject gives their peer reviewers. One major objection that can be raised against back-evaluation is that authors will reciprocate and punish reviewers who give fair but harsh reviews. We seek to evaluate the strength and dimension of the correlation between review ratings and back-evaluation ratings. A trend of reciprocation, if found, would be useful in

evaluating the modified peer review system. The lack of such trend would provide validation for its use to enhance learning in collaborative learning environments.

7. Identification of dynamics of the effects of peers' interpersonal trust in peer mentorship or any collaborative learning environment. This dissertation demonstrates how peers' trust and distrust tendencies can affect their perception of their peers' feedback and how these perceptions vary with the roles they assume in the relationship.

1.4 Dissertation outline

The dissertation is organized into eight chapters:

Chapter one contains the introduction to the dissertation and discusses the background to the research, research issues in the area, and the approach and specific contributions of the research.

In chapter two, we review the related work to the research. We start with the literature on group and peer mentorships. We extend the discussion to the literature on collaborative learning and technology that can support collaborative learning, specifically peer review and wiki systems. We narrow the discussion down to peer review and grouping for collaborative learning.

In chapter three, we present the thesis statement. This gives direction to the rest of the dissertation, with specific mention of our research questions, how we plan to address these questions, and the detailed description of our proposed modified peer review system.

In chapters four and five, we present the analyses of our studies with university students and professionals in teaching and accounting, over the past four years, using the modified peer review system. In chapter six, we discuss the implementation and evaluation of our peer-matching algorithm. In chapter seven, we present the results of our experiment with the modified Wiki system.

In chapter eight, we present the conclusion and discussion of the contributions of this dissertation and how these contributions answer our research questions, as stated in chapter three.

CHAPTER 2

RELATED WORK

This chapter first reviews literature on mentorship systems: starting from the broad definition of mentorship to the classification of mentorship. Then, it streamlines the discussion to the related work in the area of group peer mentorship. The second part focuses on the collaborative learning theories and tools, specifically Wiki and peer review systems, which can support group peer mentorship. Finally, we present the description and limitations of the existing group formation tools.

2.1 Group Peer Mentorship

The word “Mentoring” originates from the Greek mythology, when Odysseus asked his friend, Mentor, to guide and advise his son, Telemachus, as a trusted counsellor (Edutech Wiki, 2008). The name “Mentor” has since been adopted as a word in English to denote someone who imparts wisdom and shares knowledge with a less experienced colleague to promote their professional growth (Alexa, 2008). A mentor needs to show greater experience, influence and achievement than the mentee (Jacobi, 1991). Mentorship is a relationship between an older and more experienced person, called mentor, and a younger less experienced person, called mentee, focussed on the purpose of developing the mentee's career. Mentoring is not all about teaching and should not be confused with just giving advice and feedback, but a life altering relationship that triggers mutual growth, learning and development. A mentor provides two major functions, career and psychosocial (Ragins and Kram, 2007):

1. Career functions entail helping and preparing mentees to attain career advancement.
2. Psychosocial functions entail building trust, intimacy and interpersonal bonds that can aid mentees' career advancement by enhancing self-worth, professional and personal growth as well as self-efficacy of both mentors and mentees.

2.1.1 Stages in Mentoring Process

Hay (1995) defined four stages of mentoring, which can be likened to human marital relationship. These stages are:

Stage 1: “Initiation, orientation or courtship stage”: This is the stage when mentors and mentees are being introduced to one another. It starts with the initial registration, identification of goals and expectations, subsequent match-making with suitable mentors or mentees and the final agreement between the two parties to carry on.

Stage 2: “Getting established, adolescence, dependency, nurturing or honeymoon stage”: This is the review stage in the relationship, because the mentee is allowed to identify areas where help is needed and the mentor creates an atmosphere for free interaction by bridging the probable communication gap that can ensue between them due to the higher professional or age status. Lewis (1996) believed that the mentors and mentees need to move from plans to real outcomes at this stage.

Stage 3: “Maturing, developing independence or autonomy stage”: At this stage, the actual mentoring is being done with the progress monitored by the parties involved. In addition, the trust and reputation of both the mentors and mentees are judged at this stage.

Stage 4: “Ending, termination or divorce”: Termination is reached when one or both of the parties have achieved their aims, there is lack of bonding or the relationship is not fulfilling the intended aim, there is an abuse of the relationship by either of the parties or any other unavoidable need for either of the parties to leave (McKimm *et al.*, 2007; Lewis, 1996; Hay, 1995).

Kram (1983) defined the four phases of mentoring relationship as initiation, cultivation, separation and re-definition. The third stage in Hay’s definition refers to the cultivation phase of Kram’s. Also, the fourth stage of Hay’s definition covers the separation and re-definition phases of Kram’s. Bruce Tuckman (1965) also developed a four-stage group formation model, which he extended with a fifth stage after a decade. He explained that as group matures, they establish relationship and leaders change their leadership styles with time. The group formed can eventually produce a leader. The stages in his model are forming, storming, norming, performing and adjourning.

1. Forming: At this stage, group members are highly dependent on their leader and the leader must be prepared to answer so many questions. The roles of individual members

are not well known at this stage as they are just getting to know one another. So, they look onto their leader for direction.

2. **Storming:** This is when each of the group members vies for recognition and leadership. This stage might be chaotic because it involves formation of cliques and factions, as well as power struggles. It becomes very important for the group members to be focused on their goals, because their differences could cause distractions and deviations from their initial goals. However, the leader is still responsible for coaching group members at this stage.
3. **Norming:** Once the internal structure is sorted out, the purpose of the group is well defined. At this stage, the leader facilitates and enables as the group members agree and reach consensus, with their roles and responsibilities clearly understood and accepted.
4. **Performing:** Once each group member understands what they are doing and why they are doing it, they are able to perform their roles with little or no support from the leader. Although disagreements may occur within the group, members are more resolved to make positive changes than what could be obtained at the second stage (storming). Leader delegates and oversees the activities of group members at the stage, with the group members having some degree of autonomy.
5. **Adjourning:** This represents the group's break-up stage, when task has been hopefully completed, purpose fulfilled, and group members can move on to new things. This stage is sometimes referred as "deforming and mourning" because it is believed that group members would have been closely bonded and feeling of insecurity could arise from the group break-up (Businessballs.com, 2012; Chimaera Consulting Limited, 2001).

2.1.2 Classification of Mentorship Systems

Mentorship systems can be classified based on the mode of delivery and based on the number of participants and model.

1. **Face-to-face mentorship system:** This type of mentorship is characterized with face-to-face interaction between the mentors and their mentees. In this mentorship type, both sides can build a good social relationship that can enhance the success of the mentoring process.

However, it is costly and time consuming. Also, suitable mentors are not always available in the geographical location of the mentees.

2. **Online mentorship system:** This is a type of mentorship system that implements online tools to support the interaction between mentor and mentee. Participants communicate online and are not bound by geographical boundaries. However, the participants need to be trained in the use of the ICT tools and written communication can be misinterpreted (Ensher *et al.*, 2003). Online mentorship has the following advantages over the traditional face-to-face mentoring:
 - a. **Flexibility in communication:** Online mentorship with asynchronous medium of communication enhances flexibility in the time when both mentors and mentees have to send their responses to one another. In the process, participants are able to give more thoughtful and useful responses (Wong and Premkumar, 2007).
 - b. **Geographical considerations:** With appropriate technology, mentors and mentees are not bounded by their location. Mentees do not have to travel to have access to suitable mentors and they can further their interaction anywhere in the world (Stewart and McLoughlin, 2007).
 - c. **Time considerations and convenience:** Online mentoring allows mentors and mentees not to be limited by time zones by providing opportunities to interact, synchronously and asynchronously, at any time. Also, mentors are able to cover many mentees, in a group, which saves time required to do the same thing in a traditional face-to-face mentoring. With the wide spread of ubiquitous technology, mentors and mentees are able to collaborate or send messages via PC, mobiles like tablets and mobile phones, at their convenience (Wong and Premkumar, 2007; Stewart and McLoughlin, 2007; Ensher *et al.*, 2003; Sinclair, 2003).
 - d. **Cost effectiveness:** Although the initial cost of getting PC, mobiles, Internet connection and other accessories to engage in online mentoring is high, it is cheap in the long run. It allows many messages to be sent at once, at a reduced cost. Also, both mentors can reach out to many mentees at a more reduced cost than when face-to-face mentoring is attempted (Wong and Premkumar, 2007).
 - e. **Reduction in bias and generational gap:** Nationality, age and gender are factors that can lead to bias treatment in physical relationship. Since online mentorship typically

starts with a virtual relationship between mentors and mentees, it helps to break down or lessen the barriers that might lead to biased treatment. Also, online mentorship makes it easier for mentees that have a tendency to be intimidated by the profile of the more knowledgeable mentors in real life, to have open and free communication with the mentors virtually since they can assume anonymous status or use pseudonyms. The young generation use Internet more often and communicate online. Therefore, online mentoring will make it easy for them to consider being a mentor (Wong and Premkumar, 2007; Sinclair, 2003).

- f. **Improved access:** Online mentoring offers flexible access to both mentors and mentees. Therefore, it affords people with disability and mobility issues, as well as those with home and work obligations to get involved at their own convenience. 'Face-to-face meeting' between mentors and mentees can take place virtually using the VOIP technology like Skype (Wong and Premkumar, 2007; Stewart and McLoughlin, 2007).
- g. **Ecological considerations:** Face-to-face mentorship requires some travel and if in case it has to be regulated, some paper works. However, online mentoring is environmentally friendly because it minimizes travel and use of paper (Wong and Premkumar, 2007).

Some of the existing models of mentorship, based on the number and nature of the participants, are as follows:

1. **Dyad mentoring:** This is the pairing of a mentor to a mentee based on their specified goals and preferences.
2. **Triad mentoring:** This is a mentoring relationship involving three people – a senior mentor, a junior mentor and the mentee. Example is in an academic environment where a professor delegates a PhD student to supervise undergraduate or even masters' student.
3. **Peer mentoring:** This is a reciprocal peer support and learning situation, whereby peer mentors help their peers to achieve their academic, career or psychosocial goals, and they all eventually learn transferrable skills. An example of peer mentoring can take place in a workplace environment, e.g. the PHelpS system where prison guards can help each other follow procedures by filling the right forms in a new prison information system (Collins *et al.*, 1997), or in an academic environment where a more knowledgeable student

mentors less knowledgeable ones, e.g. iHELP used at the University of Saskatchewan (Greer *et al.*, 1998).

4. **Executive mentoring:** This is a top-down model of mentoring used within an organization to transfer skills and knowledge from senior staff members to their subordinates, in order to educate them and to avoid losing valuable know how when the more experienced or senior staff members retire or leave.
5. **Training-based mentoring:** This model of mentoring is targeted at specific subject of training and does not cover a broad spectrum in mentoring. For example, mentoring people on project management.
6. **Resource-based mentoring:** Here, there is no mentor-mentee matchmaking. Mentors give their details and preferences in a database, where prospective mentees can decide to choose and contact the mentors.
7. **Triangulated or Group mentoring:** In group mentoring, more than two mentees are being mentored at a time (Management Mentors, 2012; Ralph and Walker, 2011a).

2.1.3 Peer Mentoring

In traditional mentoring, mentors have to be older and more experienced (Alexa, 2008). Therefore, mentoring has been conceived to be effective only when there are experienced and older mentors. However, research had shown that peer mentoring is the strongest source of students' *cognitive* and *affective* development (Astin, 1996). In peer mentoring, participants are close in age and academic or career achievement, so they will feel more comfortable to ask for help whenever in need, than when they are paired with senior mentors (Pompper and Adams, 2006). In addition, peer mentoring helps to enhance peers' performances, productivity and also aids the development of leadership qualities in peers. However, peer mentoring is also susceptible to the following issues (Pompper and Adams, 2006).

1. unsupervised rivalry (competition) among participants, which might defeat the purpose of peer mentoring
2. lack of group cohesion among participants might also discourage learning.

Typology of Peer Mentoring

Clark *et al.* (2011) described eight types of peer mentoring, which are mostly applicable in education.

1. **Pre-entry peer mentoring:** This is a type of peer mentoring mostly offered virtually via email and social media, and allow for participants to opt in and opt out at will. For example, in higher institutions, pre-entry peer mentoring offers virtual support to mentees (new students) by linking them up with the existing students, to help their curiosity and answer their questions, before they come to the school in person.
2. **One-to-one peer mentoring at transition:** This type of peer mentoring is offered to individuals based on their needs, and they are free to opt-out when they feel that they already have smooth transition. However, it is resource-intensive because peers are paired in one-to-one matching for mentoring and to aid transition into an organization, based on the similarities of their preferences.
3. **One-to-group peer mentoring at transition:** This form of mentoring is similar to the one-to-one peer mentoring at transition, except that peers are matched to a peer mentor in groups. That is, four or more mentees may be matched to a peer mentor at once, in order to help with their transition into the organization.
4. **One-to-group transition+ peer mentoring:** This type of peer mentoring offers support not just for the transition, but also throughout the period of stay in the organization. Peers are carefully matched to their peer mentors, considering the possible duration of the mentoring relationship, all things being equal.
5. **One-to-one long-term peer mentoring:** This is a long-term one-to-one peer mentoring, which can be offered in an informal setting depending on the parties involved. It is also resource-intensive due to the one-to-one matching of peers and the fact that it involves, as mentors, peers who are more experienced and knowledgeable.
6. **One-to-group long-term peer mentoring:** This type of peer mentoring is less resource-intensive because peers are mentored in groups and can be done with, in a year or more. However, some support is required for peers to ensure the smooth running of the mentoring relationship.

7. **Partnership-led peer mentoring:** This type requires two peer mentors partnering to mentor group of mentees. For example, mentoring to aid cultural transition to another organization might require someone experienced in the old culture and another peer from the receiving culture. It might be short or long-termed, depending on how quickly the mentees can adapt.
8. **Group peer mentoring:** Group peer mentoring brings together a group of peers, who work together to mentor and help each other achieve their goals. This type of mentoring is believed to save time and cost, but it requires extra care to ensure group cohesion and reciprocity, so that the more advanced/knowledgeable peers do not feel used up while they are gaining nothing from the relationship. That is, the relationship should be symbiotic and peers should be discouraged from competing with each other, but encouraged to collaborate to help one another.

2.1.4 Factors that can affect mentoring relationship

Research in the area of mentorship had identified some factors that can affect mentoring relationship (Hodges, 2009; Thomas, 2011; Martin and Sifers, 2012). For example, Hodges (2009), in their case study research on the mentoring relationship between a nursing student and a senior staff nurse, identified lack of communication, lack of appreciation of everyday life circumstances that affect each person, differing expectations between mentors and mentees, and lack of trust as the factors that can badly influence a mentoring relationship. Also, Thomas (2011), in their article on leadership mentoring, mentioned reciprocal trust, value and commitment as the three factors that can enhance the benefits of mentorship to the participants. Straus, Johnson, Marquez and Feldman (2013), in their study of 54 faculty members from the United States and Canada, also found out that trust, respect, commitment and communication are very important for a successful mentorship. In these studies, the common factors discussed are commitment, communication and trust. There are many types of trust discussed in literature (Rotter, 1971; Wang and Vassileva, 2007). The emphasis here is interpersonal trust, which Mayer, Davis and Schoorman (1995) defined as the “*willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that party*”. There are three parts to this definition of trust – willingness to be vulnerable, expectation and risk-taking because of the uncertainty of the ability to monitor or control (McAllister, 1995). With this definition, it is easy to think that people that are high in interpersonal trust can be gullible.

However, research had debunked this assumed relationship between trust and gullibility (Rotter, 1980; Yamagishi and Kikuchi, 1999). Rotter (1980) described trust as “*believing others in the absence of clear-cut reasons to disbelieve*”. Also, Yamagishi and Kikuchi (1999) discovered that trustful people are more vigilant and detailed on the information they receive and process about other people’s trustworthiness. Therefore, they have the tendency to have a high expectation of other people’s trustworthiness.

2.2 Computer Supported Collaborative Learning

We see group peer mentorship as a form of collaborative learning, where peers with mutual accountability and responsibility work together to achieve their goals using the concepts of social interaction and collaboration. Therefore, we see a need to review the area of computer supported collaborative learning, focusing on some of the tools and theories of collaboration that can aid the success of group peer mentorship.

Computer Supported Collaborative Learning (CSCL) represents a shift from the traditional teacher-centered learning to how people can learn together in group with the help of computers (Logan, 2001; Engelbrecht, 2003; MacDonald *et al.*, 2001). Collaborative learning should not be confused with cooperative learning. Cooperative learning research began long before CSCL, and they both differ in concepts and objectives. In cooperative learning, group members split work among themselves and they assemble the parts together but they work together as one in collaboration. Therefore, cooperative learning can be said to take place at individual level, which can be aggregated when the individual group members come together. On the other hand, collaboration is a coordinated and synchronous activity, which is the result of a continued attempt to construct and maintain a shared view of a problem. Therefore, collaborative learning is a participative learning that takes place in knowledge communities, using tools of information technology (IT). It involves individuals as group members but they learn in synchrony and get involved in group interactions like negotiation and sharing (Stahl *et al.*, 2006).

2.2.1 Pedagogical Benefits of Collaborative Learning

1. *Development of critical thinking skill* (Brindley *et al.*, 2009): Since learners are just not passively receiving knowledge, but are actively involved in the acquisition of knowledge;

collaborative learning helps learners to develop strong critical thinking skill. In addition, learners are able to enhance their critical thinking because in collaborative learning (specifically online), learners retrieve information from self, others and machines, then collaborate to apply the retrieved information to current situations.

2. *Co-creation of knowledge and meaning* (Brindley *et al.*, 2009): Collaborative learning does not rely on just individual knowledge acquisition, but it provides opportunities to interact and engage in problem solving with peers, which facilitate co-creation of knowledge by community of peers.
3. *Reflection*: According to Lunsford (1991), collaborative pedagogy provides a shift from viewing knowledge as exterior and readily available, to viewing knowledge as mediated by interaction with the ecologies of collaboration (self, others and machines) (Brindley *et al.*, 2009) to generate facts and the opportunity to reproduce these facts for the creation of knowledge, a process that facilitates reflection.
4. *Transformative learning*: According to Lunsford (1991), collaboration encourages active learning, where learners engage with their peers and they also have to interpret their findings to these peers, a process similar to the occurrence in transformative learning theory (Palloff and Pratt, 2005). Transformative learning theory explains *the learning process of constructing and appropriating new and revised interpretations of the meaning of an experience in the world* (Taylor, 2008). That is, learners make meaning and learn from their experience of interaction with the world.

2.2.2 Theories of Collaboration

In this section, we have described some of the learning theories in support of the concept of CSCL. We discuss three old learning theories - socio-constructivists, socio-cultural and shared cognition theories (Kumar, 1996); and four modern theories (situated learning theory, discovery learning, connectivism and problem-based learning) that also found their roots in the old theories discussed.

1. **Socio-constructivist theory (Jean Piaget)**: Jean Piaget considered the interaction among children a powerful source of learning. Therefore, he opposed the predominant view about knowledge transmission from adult to children as a model of cognitive development.

According to Piaget, children of school age are egocentric and not able to easily bear another's point of view. However, what a child needs is something that distorts his or her "*centration*". For example, when children come together to solve a given task, exposure to the ideas of their peers who think differently affects their ego. The major goal here is to overcome this obstacle, resolve the conflicting responses and move towards more advanced form of cognitive functioning (Long *et al.*, 2011). Therefore, the distortion of their centration can eventually result in the construction of new conceptual structures and understanding (Lipponen, 2002). However, this is only possible when the relationship between the children is symmetric. That is, they are within the same age bracket. For example, when children are confronted with adults' viewpoints, they either disregard or submit due to the asymmetric relationship, and no collaboration is involved in such scenario. Putting this in the words of Piaget, "*criticism is born out of discussion and discussion is only possible among peers*" (Piaget, 1928). Also according to Piaget's view, new knowledge comes from the individual's mind and it can be shared in social interaction. However, sharing new knowledge in a social interaction leads to co-construction of knowledge through peers' ability to take others' perspectives into account (Lipponen, 2002).

2. **Socio-cultural theory (Lev Vygotsky):** According to Vygotsky, human development is a socio-genetic process through which children learn culture by interacting with their environment. However, this interaction is believed to influence children's development if it occurs within their zone of proximal development. A child's zone of proximal development is defined as "*the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance, or in collaboration with more capable peers*" (Vygotsky, 1978). Vygotsky's thoughts can be linked to collaborative learning in two ways. First, collaboration enhances the individuals' cognitive development. That is, when people engage in collaborative activities, they are able to master what they could not do before the collaboration. Second, collaboration enhances mutual engagement and co-construction of knowledge. That is, social interaction facilitates knowledge construction; which is in turn disseminated among the members of the social community (Lipponen, 2002).
3. **Shared cognition theory:** This theory is slightly different from the other two theories because it focuses on the environment in which learning takes place, whereas the others

focus on the cognitive process of learning. According to the shared cognitive theory, learners get to know the conditions under which certain knowledge can be acquired, and this helps to spur their creativity (Kumar, 1996).

Modern learning theories

1. **Situated learning theory:** Situated learning theory was developed by Jean Lave and Etienne Wenger in the 1990s. According to this theory, learning is unintentional, but it is situated with activity, context and culture. That is, learners learn by creating meaning from their interaction with the informal but contextual environment (Lave and Wenger, 1991). For example, in a field trip, learners learn by being engaged in some unfamiliar and possibly interesting environment. Also, Situated learning theory is related to the Vygotsky's theory that says that learners create knowledge by interacting actively with their environment. The six critical components of situated learning model are apprenticeship to provide guidance, collaboration to enhance the co-construction of knowledge, reflection to enable abstractions to be formed, coaching at critical times when learners are unable to complete their tasks, multiple practice and articulation in order to make their acquired knowledge explicit (McLehan, 1991; Herrington and Oliver, 1995).
2. **Problem-based learning (PBL):** According to the proponents of problem-based learning, learning is driven by challenging, open-ended problems with no wrong or right answer. Problem-based learning is a learner-centered learning model, where learners learn by engaging in real problem solving. Although some researchers had described different features of PBL (Engel, 1991; Savin-Baden, 2000), Newman (2005) described the five key (generic across all literature) features of PBL as:
 - a) **Teacher as facilitator:** The teacher is believed to be someone who is more knowledgeable than the learners. In PBL, the teacher is expected to employ their knowledge to facilitate learners' cognitive development and / or enculturation (Newman, 2005).
 - b) **Use of explicit process to facilitate learning:** The explicit process in PBL is also referred to as the tutorial process, which is used to identify the knowledge gaps in learners and the learning strategies to close the knowledge gaps (Newman, 2005).
 - c) **Use of problems to stimulate, contextualize and integrate learning:** In PBL, 'problems' refers to the materials used to initiate the learning cycle in the learners.

These problems serve the purposes of activating prior knowledge in learners, stimulating their interest in learning and setting context for knowledge acquired and their usefulness in the future (Schmidt and Moust, 2000).

- d) **Learning in small groups:** Research has shown that small groups comprising 5 to 10 members are ideal to foster collaboration among learners and also for the development of skills, and knowledge (Myers and Distlehorst, 2000; Benson *et al.*, 2001). Therefore, small groups are used in PBL to achieve the desirable learning outcomes.
- e) **Assessment and PBL:** Some of the assessment methods used to measure the learning developed by PBL program are multiple choice questions (MCQ) and modified essay questions (MEQ) (Paivin *et al.*, 1979; Knox, 1983; Swanson *et al.*, 1991).

3. **Discovery learning:** Discovery learning theory was developed by Bruner (1967) and it is based on the theories of learning by Jean Piaget, Lev Vygotsky and John Dewey (Castronova, 2002). According to the discovery learning theory, learners should discover facts and knowledge by themselves. That is, from their practical experiences, learners are encouraged to ask questions and form their tentative answers to the questions. According to Bruner (1967), there are three stages of learning - enactive, iconic and symbolic representation.

- a. Enactive: At the enactive stage, learners have a direct manipulation of objects with no deep understanding or internal representation of the objects.
- b. Iconic: At this stage learners create a mental image of external objects
- c. Symbolic representation: At this stage, learners get to understand what the mental image of objects created in the iconic stage actually stands for.

The followings are the five architectures of discovery learning discussed in literature (Castronova, 2002).

- a) case-based learning that engages students in discussion for solving case study problems (Schank and Cleary, 1994),
- b) incidental learning that takes place without knowing that learning is involved e.g. learning through game show (Bicknell-Holmes and Hoffman, 2000),

- c) learning by exploring in which learners learn while pursuing their interests e.g. use of questions and answers to spur thinking in learners (Bicknell-Holmes and Hoffman, 2000),
 - d) learning by reflection when learners try to connect their experiences with their past knowledge (Schank and Cleary, 1994) and
 - e) Simulation-based learning that involves the use of model of reality to replicate the real world in an interactive fashion e.g. use of role play (Lateef, 2010).
4. **Connectivism:** Connectivism is a modern learning theory that postulates, similarly to Vygotsky's and the social constructivism theories that learning occurs through learners' interaction with the environment, social activities and collaboration. However, connectivism theory addresses further learning that takes place in non-human materials. That is, the use of technology - computers, networks and Internet; to engage learners to acquire knowledge. Learning is acquired from a network of nodes and learners learn by being able to filter relevant nodes from the connection. This theory has been criticised as focussing on pedagogy (Verhagen, 2006) and not a new learning theory, which is not even necessary (Ally, 2008; Kop and Hill, 2008). However, Siemens maintained that connectivism is a new learning theory with the following reasons (Whyte, 2011):
- a. The existing learning theories fail to address the shift in paradigms in society, knowledge and technology. So, there is a need for a learning theory that addresses these issues.
 - b. Technology provides opportunity for *externalization of thoughts*, which is important for knowledge construction.
 - c. Technology has changed the form of human communication and subsequently altered learning
 - d. Learning should be viewed as an all-encompassing process involving *cognition, memory, emotion, beliefs and perception*. Therefore, it should not be viewed as just an acquisition of knowledge.

In summary, these theories provide a theoretical background for group peer mentorship. We see peers interacting in groups as an effective means of transferring knowledge, which, according to Jean Piaget (1928), can only take place when their peers “*distort their centration*”, instead of always relying on a senior or more knowledgeable person to mentor them. In addition,

having peers with varying capabilities grouped together can help peers to learn from their zone of proximal development (Vygotsky, 1978). In relation to the modern theories, we believe that group peer mentorship can be enhanced when peers are given problems to solve and when they engage in collaborative activities, which can also be supported by technology. Therefore, the theories of collaboration described in this section will be used to guide the conduct of our experiment and interpretation of the experiment's results.

2.2.3 Technology Support for Collaborative Learning

With the advent of web 2.0 technology, there exist many technological tools that can be used to support collaborative learning in and outside the classroom, within the six areas described below (Deal, 2009; Crow, 2010).

1. **Communication:** In collaborative learning, effective communication is very crucial to the success of the collaboration. Some of the online tools that have facilities to support communication in collaborative learning are - virtual boards, blogs, wikis / discussion boards and audio / video conferencing tools.
2. **Project management:** This comprises collaboration tools like calendar tool, time tracker, which can help with logistics like planning, scheduling and management of tasks. One major importance of these tools is to keep learners in track with the collaborative work. An example is the ENFI groupware, which gives structure to tasks for learners (Lipponen, 2002).
3. **Resource management:** Collaboration tools in this category help to handle activities storage, version control (audit), tagging and bookmarking in collaborative work (writing, project e.t.c.). Existing collaboration tools like wikis, blogs, and discussion boards have provision for these features.
4. **Co-creation and ideation:** Co-creation and ideation enhance interaction between collaborating learners in order to achieve their desired goal(s). For example, tools like wikis and peer review system enable learners to collaboratively contribute and edit their given tasks.
5. **Consensus building:** Collaboration tools make it possible for learners to engage in activities like online poll / survey in order to reach consensus on their collaborative task. Consensus

tools specifically help learners to narrow down on the ideas generated from the co-creation and ideation tools.

6. **Presentation and archiving:** Learners might need to present their feedback from the collaborative task. At the same time, the archiving tool will help to keep the log or archive of the collaborative tasks for future reference.

Since collaborative tools can affect the level of participation of learners in collaborative learning activity, it is important to consider the ease of use in the choice of collaborative tools for learning purpose (Lehtinen *et al.*, 1999). We see untapped potential in using the mechanism of some collaborative learning tools like peer-review and wiki systems, with some modifications, to support our mentorship model. Peer review is a process whereby an author's scholarly work is subjected to scrutiny by peers, who ideally are equally or more knowledgeable in the field (Smith, 2006). It is being used as a quality control tool in the research community. *Wikis are web pages that enable users to add, modify or delete contents, in collaboration with other users, typically called editors* (Wikipedia, 2014). We have discussed in sections 2.3 and 2.4 respectively, the review of related work on peer review and wiki in collaborative / group learning.

2.3 Peer review in group learning

Since its inception in the 18th century, peer review has become a veritable means of judging the quality of a product or an entity by a community of peers (Kronick, 1990; Wagner and Steinzor, 2006; Ranalli, 2011). Researchers believed that peer review system gives a sense of control to their community, and provides feedback that helps improve the quality of published papers (Chubin and Hackett, 1990; Hirschauer, 2010). Also, peer review helps in mentoring researchers, as authors, to further develop their work and knowledge by receiving competent peer-criticism (Houry *et al.*, 2012). It also helps them develop their ability, as reviewers, to provide fair and constructive criticism of peer's work by seeing the other reviews of the same paper that they have reviewed.

Traditionally, the peer review process separates the participants in the role of authors on one side from other participants in the role of reviewers on the other side (Spier, 2002). When authors submit their papers for review, the papers are assigned to reviewers by the programme

committee based on their expertise, considering their level of experience. Once reviewers have completed reviewing the submissions, authors have access to the reviews of their papers, accompanied with the decision made by the program chairs to either accept or reject their papers. In some situations, authors might be given the opportunity to write a rebuttal, to address some criticisms in the reviews, to strengthen their case for the acceptance of their papers, before the final decisions are made. In the past, authors had to send their manuscripts by mail to the conference organizers or journal editors, who, after assigning reviewers also sent the manuscripts to them for review by mail (Rockwell, 2006). The reviews were sent by mail back to the editor and from there – to the authors. Therefore the process took very long time. Currently the process is supported by web-based manuscript management systems, which enable paper submission, review assignment and the subsequent reviews to be done very efficiently. Popular existing manuscript management systems include EasyChair, Precision Conference and OpenConf, Manuscript Central, among others.

2.3.1 Learning domains where peer review can be used

In higher education, peer review has been used by teachers and learners alike. See table 2.1 for other learning domains where peer review can be used (Adewoyin and Vassileva, 2013). For example, peer review of teachers has been used in the United States school system since the 1980s at the elementary and secondary education levels, to avoid "sink or swim experience" for new teachers and generally to improve teachers' competence (Gutknecht-Gmeiner, 2005). Due to the concern about the quality of education, the peer assistance and review (PAR) program was organized to mentor new teachers and to help veteran teachers having problem with their teaching improve their competence. In this program, veteran teachers, who are in good standing, are the peer reviewers (called consulting teachers). The consulting teachers observe the classroom performances of and also counsel the participating teacher(s); which form their recommendation to the PAR panel whether to retain or layoff the participating teacher(s). The use of peer review of teachers was also introduced to the American higher institutions in the 1990s for reasons like the insufficiency of students' evaluation of their teachers and the belief that collaboration among higher education teachers is essential for educational improvement (Hutchings, 1994). Fernandez and Yu (2007) also discussed some of the benefits of peer review of teachings as a reinforcement of students' evaluation in judging the course accuracy and effectiveness of the teaching strategies implemented by the teacher. However, they raised a

caution that the peer reviewers should comprise responsible faculty, in order for the peer review process to be effective (Fernandez and Yu, 2007).

Table 2.1: Learning domains where peer review systems can be used

Goal	Mentors	Submissions	Reviews	Learning Domains
Improving research and criticism skills (Smith, 2007)	Peers, senior researchers, (program committee members)	Research papers	Peer-reviews of research papers	Academic research
Improving argumentation and writing skills (Lundstrom and Baker, 2009; Adewoyin and Vassileva, 2014)	Peer students	Essays on given topics	Peer-reviews of essays	University / High school learning in Literature, Social Sciences, Philosophy, Ethics
Improving programming skills (Hamalainen <i>et al.</i> , 2009)	Peer students	Documented source code solutions to a programming assignment	Peer-reviews of submitted program source code	Programming, Software Engineering
Seeking career advice (Gutknecht-Gmeiner, 2005)	Mentors, peers	Questions or requests for advice with situation descriptions	Advice from mentors and/or peers	Professional development, human resources

Peer review has also been used to help students improve their writing, critic and programming skills (Nilson, 2003; Hanson, 2006; Cho *et al.*, 2006; Cho and Schunn, 2007; Hamalainen *et al.*, 2009). According to Gayle Morres Sweetland centre for writing at the university of Michigan, peer review can help students develop ability to give constructive feedback and also help them to learn good writing skills from the different feedback provided by their peers on their writings (Lundstrom and Baker, 2009). The Gayle Morres centre described three strategies for implementing peer review in education. Strategy 1 is called *comment letters*, which are short essays that describe the strengths and weaknesses of the articles under review and also suggestions for revision. Strategy 2 is known as the *overview and marginal comments*, which are like the inline comments used by instructors when reviewing their students writing. In this strategy, peer reviewers provide a summary of their thoughts at the end of the article and can also provide inline comments at certain points within the article. Strategy 3 involves the use of *commenting forms*, which allow the peer reviewers to give a more detail opinion about the article under review. It is believed to be the best approach since it allows reviewers to answer some

open ended questions on what the strengths and weaknesses of the articles are and provide suggestions on how these can be improved on, by the author.

2.3.2 Criticisms of peer review

Despite the wide adoption of peer review in the research community and even in teaching, it has been criticized as not being effective and efficient, due to the fact that it is susceptible to bias, inconsistencies, abuse of ideas and excessive use of reviewers' time that could have been invested in other more profitable tasks or to generate new research (Osmond, 1983; Smith, 2006; Doucer, 2008; Walker *et al.*, 2015). According to Osmond (1983), peer reviewers can be biased towards authors that have opinions that match their views, while they can be critical of a contradicting author. Also, Peter and Cici (1982) showed in their research that the reviewers can be biased against less popular authors from less-prestigious institutions. They selected twelve papers that have been published in particular journals, reworded few parts of the papers, like the title, abstract and introduction, changed the authors' names and their institutions, then they resubmitted to the same journals where they have been published. This time three out of the twelve papers were published, while eight out of the remaining nine papers were rejected not because they lacked originality but because of supposedly poor quality. This showed that there was a bias against the authors because they are not from the top and renowned institutions (Peters and Cici, 1982). This practice, according to Osmond (1983), can discourage potential great scientists from further publishing, if they are being unfairly criticised by peer reviewers.

There are two key objectives of peer review (Cavanagh, 1996; Adewoyin and Vassileva, 2014) -

1. Summative - to assess the quality of scholarly work, and
2. Formative - to provide constructive feedback in order to mentor authors to become both better researchers, and better writers.

Unfortunately, reviews will fail to achieve the formative objective if they are de-motivating to the authors. Also in classrooms, there is a tendency for peers to refer to a diverging or peripheral review as the best, if they have not been exposed to an averagely good or more than average review. According to Smith (2006) peers have the tendencies to take different views about the

strength, weaknesses and the quality of ideas expressed in a paper; which can breed inconsistencies in their judgement of whether the paper should be published or not.

2.3.3 Existing work to improve on the peer review process

Ideas on how to improve peer review process include training reviewers, the use of open review model, blinding reviewers to the identities of authors while reviewers are anonymous (double-blind review), rewarding reviewers, standardizing review procedures, providing feedback to reviewers, and using electronic reviews (Van Rooyen, et al., 1999; Smith, 2006; MacArthur, 2007). In Smith (2006), reviewers were trained for a day and were given further instructions to use while reviewing. This research found out that training reviewers make little difference in the quality of reviews and that it is particularly difficult to train old reviewers (Smith, 2006). MacArthur (2007) also suggested that students should be given some initial instruction or training on what constitute a good review, so they can provide useful feedback for their peers.

Van Rooyen *et al.* (1999) proposed open review, where authors know their reviewers' identities. In this study, two reviewers were selected for each manuscript submitted to the British medical journal (BMJ). One of these reviewers was chosen to be anonymous while the other reviewer would have their identity open to the authors. Their reviews were then sent to an editor who would judge the quality of the reviews using a *validated quality instrument*, which comprises of five items - originality, importance of research question, presentation, method, constructiveness of comments and interpretation of results; graded on a 5-point Likert scale. The reviews were also sent to another editor, who judged the quality of the reviews independently. The authors were asked to judge the quality of the reviews, without seeing the identity of the reviewers and the decision on their manuscripts. Authors were also offered an exit questionnaire, to state the significant difference in the quality of the two reviews of their manuscripts. Their results showed that open review has no significant effect on the review quality but it increases the likelihood of reviewers declining to review (Van Rooyen *et al.*, 1999; Smith, 2006).

Also, some researchers believe that there is no quality difference associated with blind reviews (Van Rooyen, *et al.*, 1998; Justice *et al.*, 1998; Baggs, *et al.*, 2008). In Van Rooyen *et al.* (1998), one of 527 manuscripts was randomly sent to two reviewers as either blinded or open, with the reviewers' identities either masked or unmasked to the authors too. The reviews were sent to two different editors to judge their quality using their quality instrument. Their results

showed that blinding or unmasking identities does not make any significant difference in the quality of the review. However, their measure of the quality of reviews was biased because it was based on the judgement of the editors only using their quality measure instrument (Van Rooyen *et al.*, 1998).

The traditional peer review procedure comprises four stages: writing, feedback by reviewers, rewriting and publishing. However, Lan *et al.* (2011) proposed a five-stage peer review procedure that includes writing, peer feedback on the writing, *feedback on the feedback*, rewriting and publishing. They suggest that peer feedback and feedback on the feedback are important to aid learning for both the reviewers and the authors. In the past, peer review process was expensive and time consuming because the exchange of manuscripts and reviews between the authors, editors and reviewers was done by regular mail (Rockwell, 2006). However, the introduction of the electronic reviews has helped to reduce the possible delay in communication between the authors, editors and the reviewers (Komiskey *et al.*, 2010; Li *et al.*, 2012). Some researchers have worked on increasing the efficiency of electronic reviews (Cho and Schunn, 2005; Cho *et al.*, 2006; Lan *et al.*, 2011).

The existing research on peer review in science mostly discusses the faults in the existing peer review processes, such as blind, double-blind and open review, focusing mostly on whether these processes support fair, unbiased and competent reviews. Yet, there is no agreed upon definition of what constitutes a high quality review, and the criterion considered so far has been the author's satisfaction in general (with the process and fairness). Only recently the helpfulness of the review in improving the author's research paper has been considered (Cho and Schunn, 2005; Cho *et al.*, 2006; Hart-Davidson *et al.*, 2010; Xiong and Litman, 2011a, b). Cho and Schunn (2005) developed SWoRD (Scaffolded Writing and Rewriting in the Discipline), which is now called Peerceptiv. It is a web-based reciprocal peer review system that supports a four-stage writing process - writing, review, back-review and re-writing. SWoRD was designed to help students with their writing and review skills, by receiving multiple feedbacks from their papers. SWoRD includes an algorithm for measuring the accuracy of reviews and authors are able to evaluate their reviews and send their feedback to the reviewers of their articles. Cho and Schunn (2005) evaluated SWoRD with 28 university students, who were divided into three classes - *single expert feedback group*, who received feedback from an expert; *single peer*

feedback group, who received feedback from a peer; and the *multiple peers feedback group*, who received feedback from multiple peers. The three groups followed the writing procedure supported in SWORD. Their results showed that the *multiple peers' feedback group* showed a significant improvement in their writing and review skills, while the *single expert group* showed the least improvement.

We discussed earlier that one of the objectives of peer review is formative, which is, how helpful the review is for the author. Many researchers have worked on finding the helpfulness of reviews to authors. For example, Xiong *et al.* (2010) used natural language processing (NLP) techniques, on previously annotated texts from SWORD, to identify problem localization in peer reviews so as to automatically predict the helpfulness of the peer reviews, separately from the author's feedback. One drawback of this study was that the specialized features used to code the texts were manually designed by two experts - writing and content experts. However, perceived helpfulness by these experts might be different if done by peer reviewers and authors. Also, Hart-Davidson *et al.* (2010) designed a system called "Eli", which was used to analyze reviewers' feedback and also compute the helpfulness score, which is the "one time helpfulness" of the reviewer, and the helpfulness index, which is the "helpfulness over time" of each reviewer. However, "Eli" is still in test phase and the helpfulness algorithm still needs to be fine-tuned (Hart-Davidson *et al.*, 2010).

2.4 Wiki in collaborative learning

Web 2.0 has not only transformed the way information is organized on the Internet, but it has also been a game changer for Internet users and learners in particular. In higher education, wikis are being used in collaborative learning and writing because of their ease of use and the availability of options for editing by different contributors with different levels of privileges (Engstom and Jewett, 2005). Although many students in higher education are familiar with social networking sites, research had shown that they still, in most cases, lack motivation to participate in wiki-based writing (Cole, 2009; Waycott *et al.*, 2013; Wichmann and Rummel, 2013). Collaborative wikis are a type of online community, and in our case, envisaged to support group peer mentorship. Although there are claims that users are eager to engage in online communities, in practice the sustainability of online communities is a major issue due to lack of contribution by users. A typical online community comprises different types of users, based on their activities

within the community - *lurker, contributor, super user, follower, broadcaster, daily user, leader and celebrity* (Rowe *et al.*, 2013). The participation of users in online communities typically follows a power law, that is, the bulk of the contributions in the community is made by just a few active members. For example, lurkers are users that do not contribute but only access the information in the community, and they constitute about 45-90% of all the community users (Nonnecke and Preece, 2000). Similarly, *2.5% of Wikipedia editors contribute 80% of the contents in Wikipedia*, according to its founder (Rafaeli and Ariel, 2008). Therefore, it is important to motivate online users to make contributions in order to enhance the sustainability of the online community.

2.4.1 Classification of wiki editors

As it is typical of an online community, wiki users have been classified by different researchers. According to Majchrzak *et al.* (2006), wiki editors can be classified as *synthesizers* and *adders*. Synthesizers are concerned with how their reputation is influenced by what they write, and how their reputation influences the wikis and other wiki users, while *adders* are just concerned with completing their tasks and do not worry about their reputation and its effects. This classification is based on the awareness of their perceived status as the sole motivating factor for all the wiki editors. Gaved *et al.* (2006) suggests a classification of wiki users, based on their editing behaviour or type of contribution, as *placeholders, completers* and *housekeepers*. *Placeholders* make many incomplete entries; *completers* are concerned with making few but complete entries, while *housekeepers* clean up all the wiki contributions by correcting the errors made by the other users.

2.4.2 How to motivate wiki participation

Some researchers have suggested that users' high motivation is essential for successful wiki collaboration in both educational and work settings (Dow *et al.*, 2012; Gears, 2012; Arazy and Gellatly, 2013). Gears (2012) suggested that both extrinsic and intrinsic motivations are required for the success of wikis. In her research, she engaged 12 professionals in the use of corporate wiki. Her results showed that different motivational techniques can be used to increase users' participation. However, the use of wikis gives a more positive outcome when users are intrinsically motivated than when they are compelled from outside or rewarded, since after the reward is achieved, users might not see the need to continue with the use of wiki. However, Dow

et al. (2012) suggested two motivational techniques - external and self-assessment. Although both assessment techniques yielded an improvement in both the quality and quantity of users' contribution, the external assessment yielded a stronger motivational effect than self-assessment (Dow *et al.*, 2012). Also Arazy and Gellatly (2013) suggested that motivation oriented towards users' promotion (*accomplishment* and *achievement*) and prevention from cyber harm (*security* and *safety*) can trigger increased users' participation within wiki groups. Hoisl *et al.* (2007) suggested three social rewarding mechanisms (implemented in *MediaWiki*) using: 1) quality and number of references, 2) rating of articles and 3) number of views on articles. The values along these three metrics can be combined in a two-step calculation process (*revision basis* and *author basis*) to find the active participants in the wikis. *Revision basis* refers to scoring every revision to an article based on the three social rewarding mechanisms, while *author basis* refers to assigning the sum of all scores accrued from every revision that an author has made to the article. In order to reward author' contributions, Hoisl *et al.* (2007) also suggested the use of *stars* and *sparklines*. These were used to reveal the final scores of every author aggregated from the quality of the revisions they have made. Another trend is instantly rewarding editors with *barnstars*, points or warning signs that correlate with the quality of their contributions (Geiger and Ribes, 2010). However, measuring the quality of contribution (participation) in wiki has always been an open research problem.

Many variables have been used to measure participation in wiki systems. A popular measure is the use of *edit count*, which is the number of characters or words that the editors contributed to the wiki. The edit count is also the basis of the wiki power law that says that "*1% of wiki editors make 55% of the wiki contributions*" (Sub *et al.*, 2009; Antin *et al.*, 2012). However, this metric gives the same credibility to both high quality and substandard contributions (Wikipedia, 2014). Another measure is the use of editors' contributions that survive revisions by the other editors or administrators (Adler *et al.*, 2008). Geiger and Halfaker (2013) suggested the use of *edit sessions*, which are the labour hours that each wiki editor puts into making contributions to the wiki articles. In order to arrive at these time figures, they made use of the wiki log data to capture the time spent working on the wiki, by setting the maximum within session time at one hour, after which the contribution is not counted as within the same session but rather between different *edit sessions*. Although a slight correlation was found between *edit session* and *edit count* metrics, *edit session* as described by Geiger and Halfaker

(2013) might result in neglecting time spent behind the scenes doing other critical wiki activities. Also, the between session threshold of one hour might incorporate non-wiki activities, thereby overestimating the labour hours.

Many approaches have been proposed for motivating contributions to the Wiki systems (Kollock, 1999; Ling *et al.*, 2005; Joyce and Kraut, 2006). For example, Rafaeli and Ariel (2008) classified motivating factors into psychological and economic. Psychologically, some users are self-motivated to contribute, while some are obliged to keep contributing as a result of self-preservation considering the fact that they have invested their time and out of the feelings that their contributions are important to the online community (Ling *et al.*, 2005; Joyce and Kraut, 2006). Although monetary reward might count as motivation in other online communities (e.g. the Google answers website), it is less of a factor in motivating contributions in wiki (Rafaeli and Ariel, 2008). Kollock (1999) stated anticipated reciprocity, increased recognition and sense of efficacy as the three motivations to contribute in online communities. In summary, wiki is considered an important collaborative learning tool. However, strong motivation is required for wiki users to contribute and many approaches have been proposed to motivate contributions to Wiki. In addition to supporting our group peer mentorship model with the modified wiki, this study will allow us to derive specific motivational factors that will encourage participation and enhance the quality of contributions to wikis, by students engaging in collaborative writing using wikis.

2.5 General Issues in Collaborative Learning

Although there are many problems with collaborative learning, discussed in the literature (Guzdial *et al.*, 2001; Guzdial *et al.*, 2002; Wang and Burton, 2010), the following three issues are of particular interest in this research work.

1. **Group matching of peers:** Although the CSCL platform provides support for peers' interaction, it is very important to do appropriate matching when grouping peers for collaborative learning to take place. Some research had been done to describe what is the best way to match peers in collaborative learning (Kuhn, 1972; Azimita, 1988; Rogoff, 1991; Manske *et al.*, 2015). According to Kuhn (1972), grouping peers that are close in cognitive levels yields better learning outcome than pairing peers that are farther apart in cognitive

levels. However, Azimita (1988) argued that peers should not be close in cognitive levels for the collaboration to be constructive. Azimita's view was supported by Rogoff (1991) and Manske *et al.* (2015). Rogoff (1991) discovered better outcomes with adult-child involvement than child-child dominance with the fact that adult can provide guided decision making for the child while skilful peer (child) might dominate the decision making and not provide appropriate help for other peers in the group. Also, Manske *et al.* (2015), in their comparison of the performances of heterogeneous and homogeneous groups in technology enhanced classroom, showed that heterogeneous groups outperformed the homogeneous groups with higher knowledge gain. Madden and Slavin (1983) also showed that the social acceptance and academic ability of academically handicapped students improved as a result of their involvement in collaborative learning (instead of competitive learning) with the normal progress peers. This is also in support of diverse group inclusion, in order to maximize the benefits of collaboration. Also according to Gayle Morres Sweetland centre, it is very important for instructors to carefully select group size, their matching, adjust pacing and provide incentives for participation; in order for student peers to participate and learn from their collaboration through peer review. Therefore, it is important to device appropriate peer grouping algorithm that can give rise to successful collaborative learning session, in our case peer review session.

2. **Peers' motivation to participate:** The most important factor to consider in collaborative learning is the willingness of the participants to collaborate. They should be aware that they are meant to work collaboratively, helping each other to learn by providing feedback, constructive criticisms and brainstorming in a group; and not to compete with their peers (Kumar, 1996; Chua *et al.*, 2009). Chesler and Chesler (2002) described one of the factors that contribute to successful mentoring as consideration for complexities that arise from the classification of mentees or peers and motivation to participate. The existing literature on peer review, as a collaborative learning tool, lacks a discussion of the factors that can motivate peer reviewers to improve the constructiveness of their reviews, or how incentives for reviewers can be included in the reviewing process.
3. **Free-rider effect:** Since learners do not have dedicated tasks assigned to them in collaborative learning, they are only expected to pick task and work in group to achieve a common goal. Therefore, it is possible that some group members either do not work at all or

do not put in their best to the group task, since they believe that other group members will do the task. This results in work overload on few members of the group while others free-ride on their efforts (Wang and Burton, 2010).

2.6 Group matching in collaborative learning

Recommender systems can be used to recommend not only items, e.g. movies, books or items to purchase, but also people to people (matchmaking). Ensuring that both parties benefit from each other, reciprocal recommender systems have been proposed in the context of matchmaking in dating websites (Pizzato *et al.*, 2010). However, ensuring generalized reciprocity is hard when the recommendation aims to select a group of users (mentors) rather than just match two users. The main difficulties are (Masthoff, 2002):

1. *Users' preferences are not necessarily linear*: It is difficult to represent users' preferences in such a way as to capture their real intention. For example, the difference between user ratings of 7 and 8 might not be the same as the difference between 6 and 7. Also, rating scales are usually coarse. For example, users might have a preference between 'strongly like' and 'like', which cannot be well represented in discrete numeric rating scale with a limited size.
2. *Transparency versus privacy*: When modeling groups, it is very important to make the reason for certain decisions transparent to users, most importantly for usability and acceptance. However, not every user will like other group members to be aware of their preferences or capabilities. For example, in recommending movies to a group of viewers, someone might prefer X-rated movies and might not like this to be known to the other group members. Also in group learning, learners need to be aware of the reasons why certain learning materials or tutors are recommended to them. Such reasons originate from the individual preferences and needs of the constituent learners. However, not every learner would be happy when other learners see their weaknesses (Masthoff, 2002).

The core problem of group recommendation is how to use the information in the individual user models to adapt to the group needs. Group modeling strategies have been inspired by the Social Choice Theory, devised for reaching group decisions from individual opinions (McLean and Hewitt, 1994).

Many research projects have addressed peer grouping for collaborative learning in the past (see table 2.2). However, these existing grouping tools or algorithms have some limitations (e.g. limited constraints, limitation with handling "*orphaned learners*" and evaluation strategies used to determine their efficiency) (Ounnas *et al.*, 2009). In this section, we present our findings from literature on what has been done regarding group matching problems in the area of peer review and collaborative learning in general. Research in the area of group formation in collaborative learning had led to the development of certain group formation tools and algorithms, which we present in table 2.2. Four approaches to group formation for collaborative learning were discussed in the literature (Ounnas *et al.*, 2007; Aggrawal *et al.*, 2014).

1. ***Randomly selected group formation***: In this approach, instructors assign learners to groups randomly.
2. ***Self-selecting group formation***: Here, learners choose their preferred groups and they can decide to form group with whomever they choose to work with.
3. ***Instructor selected group formation***: In this approach, instructors initiate the grouping of learners based on some shared interests or some criteria shared by the learners.
4. ***Stratified group formation***: In this approach, learners are classified based on their strengths, with the top (strong) learners grouped first, followed by the next on the list and the weak learners are the last to be grouped.

Craig *et al.* (2010) described two categories of group-formation software for collaborative learning - tools that allow learners to belong to overlapping groups (Inaba *et al.*, 2001; Wessner and Pfister, 2001; Strnad and Guid, 2009) and tools that aid learners to belong to non-overlapping groups that can be applied to '*equal piles*' problem as described by Greene (2001). In both categories, grouping peers can be based on certain factors, such as their learning styles, competence, interaction, affinity, creativity and other learners' preferences (Ho *et al.*, 2009; Ardaiz-Villanueva *et al.*, 2011).

Table 2.2: Existing group formation tools in collaborative learning

Tool	Algorithm(s)	Criteria	Type of group	Drawback
DIANA (Wang <i>et al.</i> , 2007)	Genetic algorithm	Psychological factors	Heterogeneous	Too many features considered might make DIANA too generic. DIANA encourages group work but does not give room for reporting peers' gains from the group work. Also, it relies on self-reported data on students' psychological features, which can be prone to error and bias. Authors are silent on the scalability of the algorithm.
TeamMaker -> CATME (Cavanaugh <i>et al.</i> , 2004)	Random, then heuristic algorithm	Instructor-defined criteria	Heterogeneous	Limited criteria for group formation. Specified criteria are not dimensional. Some random approach is involved in the initial matching of learners. Maximum number of iterations is set to 50 for the heuristic matching and not based on stability of the group composition. Instructor-focused
OmadoGenesis (Gogoulou <i>et al.</i> , 2007)	Matrix-Hete (Hete-A), k-means clustering (Homo-A) and genetic algorithms (Mixed)	Learners' characteristics	Homogeneous, heterogeneous and mixed	Provides optional random grouping approach. Enables teachers to select learners, grouping type and criteria to consider. Flexible, but requires a designated moderator among the learners in each group. Instructor-focused.
Web-based group formation tool by Christodouloupos and Papanikolaou (2007)	Fuzzy C-Means algorithm for homogeneous group and random selection algorithm for heterogeneous group	Three learners' features defined by the instructor	Homogeneous and heterogeneous	Allows the formation of overlapping groups, which might not be suitable in a learning environment that requires formative assessment (multiple grading). Allows learners to negotiate their preferred groups, which might result in disparity in groups. Random algorithm is used for heterogeneous grouping

<i>Tool</i>	<i>Algorithm(s)</i>	<i>Criteria</i>	<i>Type of group</i>	<i>Drawback</i>
Learner-helper group by Hoppe (1995)	Rule/inference based	Learners' domain knowledge	Undefined	Group is formed based on learners' needs and might end up being a one-to-one group
Opportunistic group formation (OGF) by Inaba <i>et al.</i> (2000)	Multi-agent system	The system models the learning goals for each students and employs multi-agent systems to negotiate between the agents of each student in order to form mutually beneficial group	Homogeneous	The report lacks the description of the architecture and evaluation of the system.
I-MINDS by Soh <i>et al.</i> (2006)	Multi-agent system	Performance in the previous teamwork	Undefined	Group is self-selected by learners and also constrained by their previous performance. However, how the initial grouping is done is not clearly defined.
Expert Finder by Vivacqua and Lieberman (2000)	Multi-agent system for profile matching	Learners' expertise in specific domain	Heterogeneous	The system relies on pre-generated data to classify learners as expert or novice as in recommender systems. So, it might be prone to cold start problem.
Instructor-based group formation by Redmond (2001)	Greedy algorithm	Learners' indicated preferred time slots and projects	Homogeneous	Learners are grouped based on their preferred time slots and preference is given to learners with restricted time slots.
Rule-based tool by Tobar <i>et al.</i> (2007)	Rule-based	Based on students' characteristics and following some strict rules	Undefined	Rules are rigid and also requires manual group modifications by the instructor. There was no evaluation of the performance of the tool.
Mathematical model by Graf <i>et al.</i> (2006)	Ant colony optimization	Learners' personality traits, performance in the subject and fluency in the language of instruction	Heterogeneous	The attributes used in defining each student' vector include the language fluency, course performance, group work attitude, interest for the subject, self-confidence and shyness. However, these attributes are only relevant to existing students in the

<i>Tool</i>	<i>Algorithm(s)</i>	<i>Criteria</i>	<i>Type of group</i>	<i>Drawback</i>
				collaborative space.
*Grouping algorithms in educational settings by Agrawal <i>et al.</i> (2014)	Modified clustering	Learners' abilities in the subject of interest	Heterogeneous	The grouping algorithms described in this study have significant improvements over the random and stratified grouping approaches. However, they are mostly effective for cooperative learning. The word 'leader' was used to describe strong peer and 'follower' to describe weak peer. However, these algorithms benefit followers more than leaders, because they consider having very few leaders in each group, in order to ensure they are spread across groups and also to maximize the use of their skills to help the followers.
A genetic algorithm approach for group formation in collaborative learning considering multiple student characteristics (Moreno <i>et al.</i> , 2012)	Genetic algorithm	Arbitrary number of learners' characteristics	Inter-homogeneous and intra-heterogeneous groups	This algorithm also relies on self-report data by students to judge the characteristics used in grouping them.
The Squeaky Wheel algorithm (Tanimoto, 2007)	'Squeaky wheel algorithm'	Students' ratings of their willingness to work with their peers are used to compute their mutual compatibility, which is considered in their group assignment	Undefined	The algorithm relies on the initial ratings of students' willingness to work with their peers, which can result in students creating homogeneous groups based on the existing relationships between them, and some students, who are not preferred by their peers, end up being grouped using some neutral

<i>Tool</i>	<i>Algorithm(s)</i>	<i>Criteria</i>	<i>Type of group</i>	<i>Drawback</i>
				default values
Dynamic group formation as an approach to collaborative learning support (Srba and Bielikova, 2015)	Group technology approach	The main feature of this grouping approach is the ability to use the feedback from the previous collaboration to achieve a better grouping for subsequent collaboration	Undefined	The input data to the grouping algorithm can be generated by existing user models or questionnaire, which is not a reliable source of generating data as users can be biased in their responses. If these data sources are not used, the system might also generate the initial groups randomly.
Collaborative learning team formation: a cognitive modeling perspective (Liu <i>et al.</i> , 2016)	Heuristics algorithm; and uniform k-means to get uniform clusters of dissimilar students	This approach uses a cognitive diagnosis model named SDINA, which automatically quantifies students' skill proficiency in binary value of 0 or 1, which are used to generate collaborative learning teams with dissimilar features	Heterogeneous	This approach has only been tested with simulated data and pre-defined students' data. Since the first stage involves the quantification of students' skill proficiency, this might result in cold start problem when there is no pre-defined data to work with.
Automated group decomposition program (AGDP) (Sarkar <i>et al.</i> , 2015)	k-means algorithm; degree of heterogeneity (DOH)	This algorithm uses k-means clustering to classify students into groups and uses degree of heterogeneity (DOH) to heterogenize the group. It creates groups in two stages. In the first stage, students are grouped into the homogeneous groups using certain students' attributes like their communication skill, fluency in the use of computers and their group work attribute. In	Heterogeneous	The attributes used for creating the initial homogeneous groups are fixed and cannot be specified by instructors. Also, the tool has not been tested to assign students for actual academic projects. Also, the source of the data containing students' attributes is not defined.

Tool	Algorithm(s)	Criteria	Type of group	Drawback
		the second step, students are re-grouped to form heterogeneous groups, using their knowledge of the subject.		
GroupAL algorithm (Konert <i>et al.</i> , 2014)	Peer performance index, cohort performance index and group performance index; based on some weighted distance functions	Criteria used in grouping peers depend on the context. Also, it allows the formation of both heterogeneous and homogeneous learning groups	Homogeneous, heterogeneous and mixed learning groups	This algorithm is only tested with simulated data. It has not been used in real life collaborative learning environment.
Using constraint satisfaction to aid group formation in CSCL (Balmaceda <i>et al.</i> , 2014)	Optimization	In this approach, students reports their psychological styles and preferred team roles by answering the Myer-Briggs and Mumma's questionnaires. They also provide their social networks, in terms of the relationships with their peers. These information will then be used by the group formation assistant to recommend groups for students.	Undefined, but the algorithm aims at minimizing the distance in the students' social networks.	The algorithm relies on self-report of students' psychological, preferred role and social network data.
A deterministic crowding evolutionary algorithm to form learning teams in a collaborative learning context	Evolutionary algorithms	Students are assigned roles from the nine roles defined by Belbin. Students are then randomly placed in different groups. These groups are then reshuffled using mutation and	Heterogeneous	Still employs some random matching in the initial grouping. Also, the report is silenced about how the students' roles are determined for the initial grouping.

<i>Tool</i>	<i>Algorithm(s)</i>	<i>Criteria</i>	<i>Type of group</i>	<i>Drawback</i>
(Yannibelli and Amandi, 2012)		crossover algorithms. The fitness level of every resulting group is evaluated until the balance level is maximized for the different combination of roles in each group.		
User behaviour-driven group formation through case-based reasoning and clustering (Cocea and Magoulas, 2012)	Case-based reasoning and clustering	In this approach, learners' behaviors are modeled by case-based reasoning. Learners' strategies are retrieved from their models to derive the strategies-learners' matrix, which is passed to the clustering algorithm to group learners.	Homogeneous	This approach implements a novel idea by considering learner behavior, specifically their learning strategies, in the group formation. However, it does not address the benefits that learners derive from the groups that they are assigned to.
Heuristic Tabu Search algorithm (Hubscher, 2010)	Heuristic algorithm	This algorithm randomly assigns students to initial groups, which could change at every iteration of the algorithm until a stopping condition is reached. A stopping condition is defined by the group assignment criteria e.g. constituting evenly skilled group when there are diverse skilled-students.	Heterogeneous	It does not guarantee finding optimal solution every time.
A framework for semantic group formation in education	Semantic web technologies and logic programming	Semantic web technologies are used to create the semantic learners' profiles (SLP) and	Undefined	This approach relies on students to provide the required information needed to form the profiles, which are used for the

<i>Tool</i>	<i>Algorithm(s)</i>	<i>Criteria</i>	<i>Type of group</i>	<i>Drawback</i>
(Ounnas <i>et al.</i> , 2009)		the relationships between them, from their profile data and some initial questions that they answer about their past course experiences and relationships. Therefore, the group generator (based on logic programming) generates groups based on the constraints selected by the instructor, using the SLP.		grouping. Also, the algorithm still lacks ability to handle incomplete information about learners.

2.7 Summary

In this chapter, we have reviewed the existing literature on mentorship systems; computer supported collaborative learning (CSCL), two of the technologies that can support CSCL (peer review and wiki systems) and group matching algorithms. We believe that the features of these CSCL technologies are necessary but not sufficient to support group peer mentorship. Therefore, in order to support our group peer mentorship model, we have proposed the solution in chapter 3 (see section 3.3).

CHAPTER 3

THESIS STATEMENT

The goal of this dissertation is to support group peer mentorship with online collaborative learning tools. We address some specific objectives while determining the best tool and approach to support group peer mentorship, and these will be discussed as part of the research questions in the following section. In this dissertation, “mentorship” and “mentoring relationship” are used interchangeably. We are also considering reviewers as mentors and authors as mentees. As a result, “authors and mentees” and “reviewers and mentors,” will be used interchangeably in this chapter and in the subsequent chapters.

3.1 Research Questions

The main research question here is: “Can online collaborative learning tools support group peer mentorship?” As discussed in the literature review (chapter 2), there are many collaborative learning tools in use. We want to find a tool that is suited for our model of mentorship with respect to the existing features of the tool and users’ adaptability to the use of the tool for the purpose of mentorship. In our preliminary findings, an online peer review system seems suitable candidate. The main research question is thus rephrased as: “Can an online peer review system support group peer mentorship?” We then divide the main research question into four separate questions:

1. Research question 1 (RQ₁): *How do we compose the review/mentoring group? Does the diversity approach used to group peers achieve successful mentoring sessions? As applied to peer-review systems, this question translates to finding an appropriate reviewer assignment mechanism.*
2. Research question 2 (RQ₂): *Are incentives required for providing high-quality feedback to peer mentors, and if yes, what incentives? Does anonymity enhance the quality of mentorship?*
3. Research question 3: (RQ₃): *What metrics can we use to measure the success of the mentoring relationship? In mentorship, it is important for a mentor to provide formative feedback to help the mentees improve in the area of need, which is the purpose of*

mentorship. How do we ascertain that this objective is fulfilled in the mentoring relationship? Because it is group peer mentorship, we believe peers in both roles, mentors and mentees, should be able to learn to improve their skills. How do we evaluate if both mentors and mentees have learned something that helps to improve their skills?

4. Research question 4 (RQ₄): *What factors can influence peers' feedback in the mentoring relationship? Do peers reciprocate their feedback to look good to their peers? How do peers' interpersonal trust scores affect their ratings behaviour both as mentors and as mentees?*

3.2 Answering the Research Questions

In this section, we discuss and justify our approach to the research questions listed in section 3.1.

1. To answer RQ₁, we implement a peer-grouping algorithm based on the traditional Hungarian algorithm. We test this algorithm manually in studies with human participants and also with system generated data. We also compare our algorithm with three of the existing grouping algorithms used for grouping learners for collaborative learning, considering the knowledge gains of peers from each group formed and the time and space consumption of the algorithm.
2. To answer RQ₂, we consider different motivation strategies that can help peers to give high-quality feedback. In the proposed approach (section 3.3), we offer intrinsic and extrinsic motivations for peers. Intrinsic motivation comprises peers who provide feedback as both mentors and mentees. We believe the feedback that the mentees give their peer mentors motivates them to improve the quality of reviews (feedback) that they give the mentees. Also as peer mentors, they will be able to see the feedback given to their mentees by other peer mentors. This will enable them to learn from other peer mentors. As a form of extrinsic motivation, we propose that peers use pseudonyms throughout the peer mentorship process to encourage them to give thorough and critical feedback; this enables peers to accrue their feedback-ratings reputation over time. The community of peers can use this rating to judge competence in the future.
3. Because we are supporting the group peer mentorship with online peer review systems, we consider the features and process of online peer review systems that can support this mentorship model. The major objectives of peer reviews are to assess the quality of scholarly

work (summative feedback) and to give constructive feedback to help authors and reviewers improve their skills (formative feedback). This feedback can be judged as constructive only by the review recipients through the back-evaluation of their reviews. Back-evaluation is the process whereby review recipients (authors) provide an evaluation of the reviews that they receive on their paper. The traditional peer review process does not include back-evaluation of reviews by the authors. Instead, only the program chair, committee, or the editorial team have the chance to evaluate the quality of the reviews and, in some extreme cases, to disqualify the reviews or the reviewers (Ranalli, 2011; Do, 2014). However, the author is in the best position to judge whether the reviews are constructive and helpful in improving their work. If authors lack the opportunity to evaluate the reviews of their papers, reviewers cannot receive the best possible feedback that could motivate them and help them improve the quality of their reviews. In other research, authors' perception and the success of the peer review process are determined by whether their papers are accepted (Gibson *et al.*, 2008; Weber *et al.*, 2008). This is one sided because the decision to accept a paper does not necessarily reflect the reviews' quality and thoroughness. To answer RQ₃, we solicit the authors' (mentees) feedback on their learning, review helpfulness, their satisfaction with peer feedback, and the evaluation of the mentorship process by peers in both mentor and mentee roles.

4. To answer the first part of RQ₄, we want to investigate whether peers are inclined to help each other using the peer review system and that they are not just reciprocating ratings to formally fulfill their part in the mentorship process. To confirm this, we measure the correlation between mentees' feedback ratings and the back-evaluation ratings that they give the feedback from their peer mentors. If the mentees are trying to reciprocate their peer mentors' good ratings and positive comments or retaliate for the mentors' low ratings and critical comments, we would expect to find a positive correlation. To answer the second part of RQ₄, we compute the trust and distrust tendencies (TDT) of peers using the Rotter interpersonal trust scale. We then use correlation analysis to determine the direction of these two relationships:
 - i. The relationship between mentors' interpersonal trust scores and the review ratings they give their mentees.

- ii. The relationship between mentees' interpersonal trust scores and the back-evaluation ratings they give their peer mentors.

3.3 Proposed Approach

We propose the use of two modified collaborative learning tools to support the group peer mentorship: peer review system and wiki system.

3.3.1 Modified Peer Review Process

We find similarities between the concept of group peer mentorship and the conference peer review procedure. Peer review is a quality control process for judging the quality of scientific contributions and for providing formative feedback to the authors, and it plays an important role in the process of educating aspiring young researchers. In the past, peer review was done via regular post, which could take months to conclude. Now online peer review conference management systems such as EasyChair and OpenConf have shortened the submission-review-feedback cycle to one to two months, which is mostly the time required by reviewers to read and evaluate the manuscript. In typical online peer review systems, users can be authors, reviewers, or both. As authors, users have access to functions that allow them to submit their papers, which are then assigned for review through the peer review system to competent reviewers, and receive the reviews of their papers, with a decision of either acceptance or rejection (for conferences and journals), and a list of recommended revisions. As reviewers, users are typically able to see the papers assigned to them and submit their evaluations and reviews. Meanwhile, some conference peer review systems provide a more advanced process, allowing reviewers to see the other reviews of their assigned papers, to discuss the papers with the other reviewers, and, eventually, to modify their own reviews before the decisions are made. Some review processes also support a rebuttal phase, giving authors a chance to respond to the criticisms raised by reviewers before the final acceptance decisions are made. In a way, this emulates the journal paper submission process. The relationship between authors and reviewers in an advanced peer review system can be likened to the relationships between mentees and mentors in a group peer mentorship system (Adewoyin and Vassileva, 2013).

Despite the wide acceptance of peer review in the research community, some researchers have criticized it as being slow, expensive, and plagued with biases and inconsistencies (Do,

2003; Smith, 2006). The existing research has mostly discussed the faults in the current peer review process. However, they have not succeeded in providing adequate and unbiased criteria that can be used to determine a high-quality review. Only over the last 10 years has the literature considered the helpfulness of the review in improving the author's research paper (Cho and Schunn, 2005; Cho *et al.*, 2006; Hart-Davidson *et al.*, 2010; Xiong and Litman, 2011a, b). In addition, the existing literature lacks a discussion of the factors that can motivate reviewers to improve the constructiveness of their reviews or how incentives for reviewers can be included in the reviewing process. Also, appropriate measures of the success of a peer review session from the points of view of both reviewers/editors and authors need to be developed. The features of a typical peer review systems are necessary but not sufficient to support group peer mentorship (Adewoyin and Vassileva, 2013).

To overcome the problems described above in order to support the group peer mentorship system using a peer review system, we propose the following review mechanism modifications:

1. Peers will be grouped based on their competence. An initial calibration test will be given to all peers to judge their competence. The test results and their profile information will constitute their user model. Groups of peers reviewing each other's works will be formed by ensuring diversity among the members. For example, we do not want peers with low competence to review one another. So, peers with high competence would be grouped with weaker peers. More competent peers should serve as mentors to their less experienced peers in the same group. Also, there will be a senior peer, e.g. a teaching assistant in the case of a university, to provide authoritative feedback on the group's performance.
2. Peers will remain anonymous in each group, by using pseudonyms, so they can build continuous identities and reputation over multiple sessions. This can be an incentive for some peers who want to build a good reputation by mentoring others. To provide an oversee and across-group standard / consistency, there will be a senior, not anonymous peer to reinforce their feedback on peer reviews in case of conflict and to monitor the overall group performance.
3. We propose that peers in the role of reviewers will get feedback from the paper authors about their reviews. We propose also that peers as reviewers should see the reviews given by other

reviewers on the same paper and provide feedback on the quality and usefulness of these other reviews, given by the other peers in their group. This should allow peers to learn how to provide high quality feedback (reviews). It should also motivate them to improve the quality of reviews they give, since their reviews are open for others to see.

- To judge the overall success of the review session, we propose that peers as reviewers provide an evaluation of their learning from other reviews and feedback from authors and other reviewers, and peers as authors evaluate the learning from all received reviews of their papers.

To implement these modifications, we have proposed the mechanism shown in Figure 3.1. We conducted experiments using the proposed modified peer review framework with two online peer review conference management systems, EasyChair and Peerceptiv (formerly called SWoRD). These experiments and the results are presented in chapters 4 and 5.

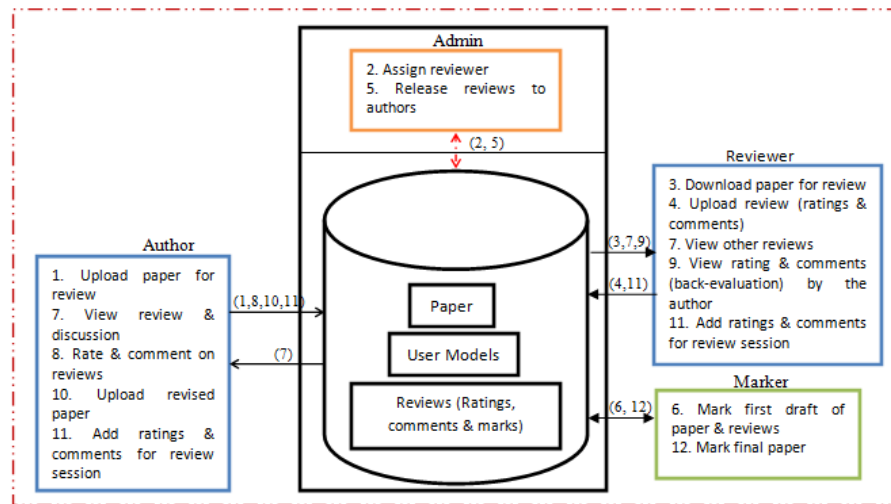


Figure 3.1: Framework of the modified peer review system

3.3.2 Modified Wiki

While exploring different collaborative learning tools that could support online group peer mentorship, we considered the alternative of open wiki-style collaboration. In this collaboration mode, everyone participates in the creation of a document, and all edits and changes are visible to everyone both in the current version of the document and in the document's history, which shows a chronological list of all the changes made and who made them. To facilitate the peer-

feedback function of wiki editing, we modified the traditional wiki editing process, adding new features to the wiki system. These include:

1. An email notification that alerts users of changes made to their contributed contents.
2. Notification interface for content changes that lists all the modifications of the current user's contributed contents and highlights these changes within the changed wiki document.
3. The interface includes an option for the user to accept/reject the change and evaluate it (rate it) and provide a written feedback / justification for the decision, or just a comment.

The notification facility serves a role to create awareness of the peer-feedback on the contributed content and to create incentive for the peers to edit the contents contributed by others because they can earn good reputation if their changes have been rated highly by the contributors.

We conducted a modified wiki experiment alongside one of the modified peer review experiments with the same set of participants (students in a University undergraduate class, Ethics and IT 2014) to determine the preferred collaborative learning tool by the users for group peer mentorship. The experiment and the results are discussed in Chapters 4 and 7 (see section 4.3.2 and section 7.4).

3.3.3 Group formation and review assignment

Peers have different skill levels. The problem in using a peer review or wiki system in group peer mentorship concerns how to group peers who provide feedback to each other (i.e. how to do the reviewer assignment or the wiki editors' combination) to ensure that all peers can benefit from the collaborative process, where everybody has something to offer. This problem is akin to the group formation problem in the area of CSCL, where some researchers support having peers that are close in skill level grouped together (Kuhn, 1972), whereas others believe that grouping a mix of peers with different levels yields a better learning outcome (Azimta, 1988; Rogoff, 1991; Manske *et al.*, 2015). A lot of research on group formation in CSCL has been influenced by the theory of Vygotsky, and specifically his "Zone of Proximal Development" concept, which is the difference between what peers can do without help and what they can do with help from a teacher or more capable peer. According to this theory, peers should be grouped with other peers with varying degrees of ability. Therefore, we propose that peers with abilities ranging from

weak to average and strong be grouped together. To ensure that all peers have the opportunity to receive valuable feedback, we propose a matching algorithm⁴ based on the principles of Hungarian algorithm for assignment problem with the following constraints.

1. Each group should not exceed four peers in number. Research had shown that a group of three to four peers works best because it is easy to coordinate and it also encourages group members to be actively involved (Davis, 1993; Csernica *et al.*, 2002). Also, because peers will be reviewing each other's work or performance and we do not want any peer to feel overwhelmed by the volume of work, it is important to keep the number of peers in each group to four so peers can handle a reasonable number of reviews at once (three in this case) with which they should be able to give their best effort.
2. We propose peers should be classified as weak peers (*wp*), average peers (*ap*) or strong peers (*sp*) by using an initial calibration task relating to what they are expected to do in the peer review. The feedback from the initial task will be used to temporarily determine their strength, which changes after each peer review iteration. To classify them as strong, average or weak peers, we will calculate the average score from the initial calibration task for all the participants; every participant with a score less than the average is classified as a *wp*, while every participant with a score equal to the group average is classified as an *ap* and every participant with a score greater than the group average is classified as an *sp*.
3. In each group, the number of *wp* should be less than half of the total number of peers in the group (i.e. $\text{number}(wp) < [\text{number}(wp+ap+sp)]/2$). This condition will help ensure that the peers form a bipartite graph, and if not, we have to eliminate the extra link to be able to apply our variant of the Hungarian algorithm to match the group of peers.
4. A weak peer cannot review/mentor another weak peer. This will ensure that peers are able to learn from other peers in their zone of proximal development (Vygotsky, 1978). That is, *wp* should get help from more knowledgeable and more skilful peers.
5. A *sp* or *ap* cannot receive more than one review (be mentored) from a *wp*. This is to ensure that every peer gets more than average and useful feedback overall. Although we would like

⁴ The algorithm is presented in Chapter 6 (see section 6.2)

every peer to be able to see and learn from an example of a not-so-good review, we still want peers to be able to learn from their good peers in order to be able to operate within their zone of proximal development (Vygotsky, 1978).

CHAPTER 4

ADAPTING ONLINE PEER REVIEW TO SUPPORT GROUP PEER MENTORSHIP

This chapter presents five case studies of using online conference peer-review systems as group peer mentorship systems. Four of the case studies were conducted with graduate and undergraduate students of Computer Science, while one case study was conducted with the professional accountants who are enrolled in the masters in professional accounting (MPAcc) programme at the Edwards school of business.

4.1 Introduction

Web systems provide opportunities for online mentorship that overcome the location- and time-constraints of traditional face-to-face mentorship. Thus, the Web has allowed for newer forms of mentorship to emerge, such as peer mentorship and group mentorship. Yet, web-based platforms for management of mentorship process are not available. In the last chapter we proposed that online peer review systems can be used to support the management of group peer mentorship. Peer reviews have been used in the academic community for controlling the quality of new research publications and also as a means of providing constructive feedback to authors, helping them to improve their research and communication skills. Some of the important issues to be addressed in the management of both peer review and group peer mentorship are:

- i. finding competent and constructive mentors for particular mentees; and
- ii. motivating mentors to give honest, critical, detailed and friendly feedback to mentees.

Group peer mentorship system is a variant of group mentorship, which is a collaborative relationship where peers are grouped based on their capabilities and they mentor each other within the groups, instead of having to selectively recruit mentors (senior mentors). Traditional mentoring presumes that mentors need to be older and more experienced. However, research had shown that peer mentoring is the strongest source of students' *cognitive* and *affective* development (Astin, 1996). We have adopted the group peer mentorship model, where students are formally equal and they mentor each other.

In this chapter, we present five case studies of using online conference peer-review systems as group peer mentorship systems. In four cases, the conference peer review systems were used

to develop writing and argumentation skills in graduate and undergraduate students of Computer Science, while in one case, the peer review system was used to develop analysis, audit and review skills in accounting professionals. The purpose of these studies is to determine how peer review system can support group peer mentorship. The following are the additional questions we want to explore in order for the peer review system to support our mentorship model.

1. **Does the diversity approach used to group peers achieve successful mentoring sessions?** Peers will be grouped based on their competence. An initial calibration test will be given to all the peers to judge their competence. The test results, in addition to their profile information, will constitute their user model. If it is not possible to conduct a calibration task, participants' past course averages can be used. Also in each group, diversity will be embraced. For example, we do not want peers with low competence to review one another. So, peers with high competence will be grouped with peers that are weaker. We believe that more competent peers will serve as mentors to their less experienced peers in the same group.
2. **Does anonymity enhance the quality of mentorship?** Peers will remain anonymous in each group, but they will have pseudonyms so they can build continuous identities over multiple sessions and reputation. For the studies in the graduate and undergraduate classes of Computer Science, there will be a senior peer (marker), who will not be anonymous in order to reinforce their feedback on peer reviews, in case of conflict, and the overall group performance.
3. **Can author's and other peer's feedback on the reviews serve as incentive for providing high quality reviews?** We propose that peers in the role of reviewers to get feedback from the authors on the reviews they give them and also see the reviews given by other reviewers on the same paper. This, we believe, will allow peers to learn from others and will also motivate them to improve on the quality of review they give. We propose also that peers as reviewers should provide feedback on the quality and usefulness of the other reviews on the same paper, given by other peers in their group. Both the author's and other reviewers' feedback will contain a numeric component, which reviewers can accrue over time with their pseudonyms and this value can be used by the community of peers to judge the competence (reputation) of peers for future engagement.

4. **What metrics can we use to measure the success of a mentoring relationship?** In order to judge the overall success of the review session, we propose that peers in the role of reviewers provide an evaluation of their learning from the other reviews of the papers that they review and from the back-evaluation ratings (feedback on the reviews that they give) that they receive from the authors, and peers as authors evaluate their learning from all received reviews.

This chapter is organized as follows: In section 4.2, we present our methodology, which comprises the mentorship workflow and settings used in EasyChair to implement the group peer mentorship system, the strategy used to form groups and match students in the roles of reviewers and authors. The results of 5 case studies are presented in section 4.3; section 4.4 concludes this chapter.

4.2 Methodology

We carried out the case studies as part of students' coursework in graduate and undergraduate classes that have strong writing component. The goal of mentoring in this case was to develop better writing and argumentation skills in the students. Students were both authors (mentees) and reviewers (mentors), which is typical of a *peer* mentorship system. The features of a standard online peer review system (e.g. EasyChair) were appropriate, but not sufficient to support online group peer mentorship in this case. Our peer-review framework implemented group peer mentorship and improved the review process to include feedback on reviews by authors (closing the peer-mentorship loop, so that mentors can also learn from the mentees) and feedback on the peer review process by both authors and reviewers (see Fig. 3.1 in section 3.3.1).

4.2.1 Implementing group peer mentorship

Information about the skills and competence of the mentors, as well as information about the goals and features of the mentees, can be used to match mentors to mentees. Since our approach is that of *group* mentorship, the students were organized to review each other's work into different reviewing groups based on their relative strength as writers and critics. We classified student as weak, average and strong peers based on the results of an initial calibration task - an essay, statement of interest, or survey; the produced documents are analyzed to establish the individual students' levels of (writing and critiquing) competence. We adopted a closed grouping of peers, based on their levels of competence and in such a way as to ensure a mix of expertise

levels that complement one another. For example, peers with very high marks in the initial calibration test were grouped with those that are somewhat weak, and peers reviewed each other's submissions. We decided to match weak peers with strong peers in order to enable the weak peers to learn from their zone of proximal development (Vygotsky, 1978). We believe that the more competent peers would serve as mentors to their weaker peers in the same group, and the less competent students would be mentees. By reviewing the papers of the mentors, they would get exposed to a supposedly better organized argument and better written documents. Also, we believed that the more competent peers would learn from seeing examples of weak essays and probably reviews. As it was not a conference, acceptance or rejection was not relevant.

4.2.2 Feedback Loop

In order to encourage group discussion and peer-learning both for mentors and mentees, the students in the role of reviewers or mentors could see the reviews that the other reviewers or mentors gave to the works assigned to them to review. Thus everyone in the role of a mentor could see and learn from the feedback provided by the other mentors after one had submitted his/her own review, to avoid bias or free-riding. A similar feature exists in some conference peer-review systems, such as EasyChair (yet it is not a default feature).

Also, the peer-review system settings should allow discussion between the author (mentee) and the peer reviewers by rating and commenting on the quality and helpfulness of the reviews they got. This is similar to a rebuttal process adopted by some conferences. In addition to the authors submitting the revised version of their work, all group participants could submit evaluations of the review process, including the discussion and feedback provided by each review.

To implement this modified group peer-review model, we followed the mechanism in Figure 3.1 (see chapter 3). Please note that “PC members” (reviewers) here are the students in mentoring role, and “Authors” are the same students but in the role of mentees. To provide a common base for discussion, weekly writing topics were assigned (common for all students), for which the students had to submit short discussion papers. The papers were evaluated through anonymous peer-review process (every author was also a reviewer of at least two other papers). As peer-reviewers (mentors), the students were instructed to provide constructive feedback. As

mentees, the student authors could take the feedback into account in preparing the final version of the paper, which after submission, should be evaluated by an expert to provide an objective measure of quality (e.g. by a teaching assistant or instructor in the class). After each reviewing cycle, authors (mentees) were asked to provide feedback to the reviewers (mentors) on the quality and the helpfulness of their reviews. Also reviewers (mentors) were asked to provide feedback on the reviewing session of each individual paper, whether they learned and improved their skills for constructive critiquing, from seeing the other reviews of the paper and from the ensuing discussion.

4.3 Case Studies, Results and Discussion

Between January 2013 and April 2016, we did five case studies of exploratory nature, in undergraduate and graduate classes of computer science and in a masters in professional accounting class (MPAcc), all which have a strong writing component. These were field studies, rather than lab studies, so many different accommodations needed to be done to fit the requirements and constraints of the particular class (e.g. coursework, number of students participating, etc.). This necessitated implementing the peer-review mechanism with different parameters, summarised in Table 4.1). The results from the five experiments, while not statistically significant, due to the small class sizes, show consistent results over time with different classes, coursework organization and types of students. They are presented in the subsequent sections.

Table 4.1: Overview of the five case studies

<i>Date</i>	<i>Class at the University of Saskatchewan</i>	<i>Number and year of participants</i>	<i>Number of peer review assignments completed during the term</i>	<i>Task</i>	<i>Anonymity / Pseudonymity</i>	<i>Peer review tool</i>
Jan - April 2013	Ethics and IT (CMPT 408)	6 participants 4 th year undergrad students	7	Weekly essay and review of each other's essay	Anonymity	EasyChair (back-evaluation using Google form)
Sept	Research	24	5	Paper	Pseudonymity	SWoRD

<i>Date</i>	<i>Class at the University of Saskatchewan</i>	<i>Number and year of participants</i>	<i>Number of peer review assignments completed during the term</i>	<i>Task</i>	<i>Anonymity / Pseudonymity</i>	<i>Peer review tool</i>
2013 - April 2014	Methods (CMPT 880/890)	participants, 1 st year graduate students		summary and review of each other's summary		(back-evaluation within the system)
Jan - April 2014	Ethics and IT (CMPT 408)	10 participants, 4 th year undergrad students	6	Six weekly essays and review of each other's essay using the peer review system; four collaborative essay writing using Wiki	Pseudonymity	SWORD (back-evaluation within the system)
Jan - April 2016	Ethics and IT (CMPT 408)	16 Participants, 4 th year undergrad students	6	Weekly essay and review of each other's essay	Pseudonymity	Peerceptiv (back-evaluation within the system)
Oct - Dec 2015	Professional Accounting Class, 2015	17 participants, Masters of professional accounting class	2	Analysis of audit cases and review of each other's analysis	Pseudonymity	Peerceptiv (back-evaluation within the system)

4.3.1 Exploratory Study - Ethics and IT, 2013

We did a small-scale case study in a fourth year undergraduate computer science on Ethics and IT, which had mostly an exploratory role. Our goal was to find out if group peer mentorship in improving students' writing and argumentation skills can be accomplished within regular coursework consisting of uniform writing assignments. The study lasted 10 weeks, of which in 6 the students had to submit a short weekly essay on a specified common topic or question, related to the content discussed during the week. These weekly writing exercises were concluded in the

end of the term with one long paper (8-10 pages) on an individual topic related to the class and chosen by the student, in consultation with the instructor. Thus there were 7 peer review sessions, all supported with EasyChair, a free online conference peer-review system. There were 6 participants, of which all were Canadian, male, age 20-30, and all but one were native English speakers. Participants were grouped for the peer review session based on their competence, in such a way that weak peers were grouped with strong peers in order to ensure diversity and also allow peers to learn within their zone of proximal development. We conducted an initial calibration test, a short essay, which was used to judge participants' competence for the first peer review session, and their subsequent competences were derived from their peer ratings from each peer review session. The students were asked to fill brief questionnaires after every peer-review assignment, and one final questionnaire at the end of the class about their overall experience (see Appendix B). We had a marker, who can be considered a senior mentor and not anonymous, in order to make her feedback more authoritative. The marker gave feedback on both the quality of the papers and the reviews. This feedback was released to peers just before they filled the weekly questionnaire and before they engaged in the next peer review session. The feedback was both formative and summative, since it contained both comments including improvement suggestions and a grade that was counted towards the total grade in the course (the latter was required to ensure that the students actually do the peer-reviews).

With this small exploratory study, we hoped to find answers to the four research questions, discussed in section 4.1, regarding the potential of our approach to support effectively group peer mentorship.

1. Does the diversity approach used to group peers achieve successful mentoring sessions?

According to our method the students were given an initial calibration task (a short essay on a given topic), which was graded. The student grades were used to group the students to peer-review each other's writing so that in each group there were both high performing students and weaker ones. In the subsequent sessions, the student grades from the previous sessions were used to group them using the same principle. We had seven review sessions. In each session, we re-grouped peers using their peer ratings from the previous review session. However, the group compositions in the first and seventh sessions happened to be the same.

To evaluate the success of each review assignment (i.e. the peer-group composition), the students were asked after each of the 7 review cycles to evaluate their satisfaction with the quality and helpfulness of the reviews they received from their peers. The mean and standard deviation of their responses over the 7 assignments are presented in table 4.2. The students tended to agree with the positive statements and seemed generally satisfied with the reviews they received from their peers. To verify the answers of the students and minimize a bias that may have been present due to the small size of the group, we asked also several negative questions, where students had to show their agreement with the reverse positive questions. Though ideally the sum of the means of the positive and negative version of the question should sum up to 10, while our results show that a small proportion of the students agreed with both the positive and the negative version of the question, the levels of agreement with the negative statements (lacking friendliness, lacking good points and lacking useful suggestions) were lower and the standard deviation - higher. This result supports the hypothesis that the choice of diversity of group members is effective in achieving a successful mentoring session.

Table 4.2: Authors' ratings of the reviews they received. The ratings are expressed as level of agreement with the statements expressing the criteria, using a scale of 1 (lowest) to 10 (highest)

<i>Criteria</i>	<i>Mean</i>	<i>Standard Dev.</i>
Good points raised	8.50	1.34
Useful corrections to style and grammar	8.33	1.63
Suggestions on good ways of expressing ideas	8.17	2.23
Friendliness	5.17	2.86
Lack friendliness	3.67	2.07
Lack good points in the review	4.00	3.03
Lack useful suggestions to style and grammar	4.00	3.69

Were the students aware that there was some principle behind the review assignment?

To find out, the students were asked in the final (exit) questionnaire, if they thought the selection of peer-reviewers was random or if they thought it was based on certain principles. The majority (71%) of the students thought the assignment was based on some principle and 29% thought it was random. Of those that thought it was non-random, some provided the following comments:

- *"I think it was based off of the initial 0 mark essay we did at the beginning of the year. I feel it was actually marked and the group split up that way to see the progression of those groups...."* (Study participant)
- *"It would be some type of analysis I would guess... based on past marks as to show understanding of certain topics or something along these vague ideas".* (Study participant)
- *"I think everybody reviewed my papers. The quality of reviews varied substantially in my opinion, so I would guess the quality was based on the reviewer."* (Study participant)

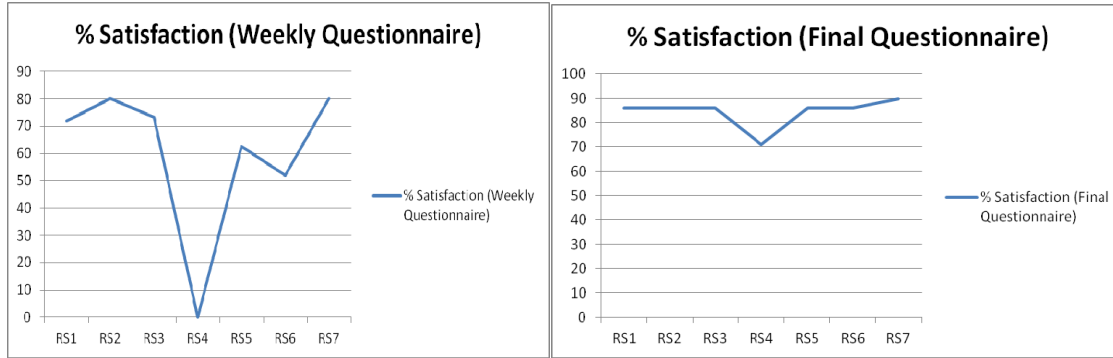
The answers seem to show that, even though the majority of the students guessed correctly that the choice of reviewers was not random, they did not understand the principle of the peer reviewer assignment, or who were their mentors or reviewers, so anonymity was preserved, despite the small size of the group.

2. Does anonymity enhance the quality of mentorship?

In six out of the seven review sessions, peers were kept anonymous. The participants had difficulty in recognizing if their reviews were done by the same reviewers in every session; the answers were perfectly split, with 43% choosing “yes”, 43% - “no” and 14% - “hard to say”. Obviously, despite the small number of participants, the anonymity worked well; as the peer reviews were only formative, there was no particular incentive for trying to game the anonymity of the system and finding out who is reviewing one’s paper. This contributed to unbiased and constructive reviews.

In the weekly questionnaires that followed each peer-review session, the authors were asked to state their satisfaction (in 100%) with the session. As can be seen in Figure 4.1a, the average satisfaction varied between 50% and 80% over the sessions. Interestingly, for the one session for which the review assignment was not anonymous (i.e. the group membership was visible to everyone), the students did not provide any satisfaction values (therefore the 0 value). The final questionnaire asked them once again to estimate their satisfaction with the seven peer-review sessions. This time the satisfaction results were higher, between 85 and 90%, and the students provided an estimate of their satisfaction also for the non-anonymous week which was the lowest among all the sessions (70%), see figure 4.1b. According to the marker’s feedback,

the reviews provided by peers during the session when their identities were revealed were more polite, less critical and not as thorough or constructive as those that the same students had written in the other sessions, when they were anonymous.



(a)

(b)

Figure 4.1: Authors' satisfaction with the reviews received in each reviewing session: (a)-on a weekly basis, (b)-in the final questionnaire.

Tables 4.3 and 4.4 give the summary of the marks given to their final essays and reviews by the marker. Although there is no significant trend in the marks of peers from week to week, we discovered that peers have the least marks for their final essays and reviews on the fourth week, when their identities were revealed to their peers.

Table 4.3: Marks given to the peers' final essay by the marker

	<i>Essay 1</i>	<i>Essay 2</i>	<i>Essay 3</i>	<i>Essay 4</i>	<i>Essay 5</i>	<i>Essay 6</i>
A	8	7	10	9	7	7
B	8	9	7	10	9	8.5
C	10	8	8	8	6	9
D	8	9	8	10	9	6.5
E	8	9	8	9	10	8.5
F	10	7	8	0	10	9
Average marks	8.67	8.17	8.17	7.67	8.5	8.08

The lower marks given by the marker to the 4th essay and to the peer reviews for this essay may be the main reason for the dissatisfaction of the students with the 4th review session. While the lower marks on the essay per say may be due to a more difficult topic (though we could not confirm this), the lower marks on the reviews is likely due to the lack of anonymity during this session.

Table 4.4: Marks given to the peers' reviews by the marker

	<i>Review 1</i>	<i>Review 2</i>	<i>Review 3</i>	<i>Review 4</i>	<i>Review 5</i>	<i>Review 6</i>
A	4	4	4	4.5	2	5
B	5	4	5	2.5	4	4.5
C	4	4	4	4.5	3.5	4.5
D	4	5	5	2.5	5	5
E	5	5	4.5	2.5	4.5	5
F	3	5	4.5	0	4.5	4
Average marks	4.17	4.5	4.5	2.75	3.92	4.67

Although the traditional face-to-face mentoring encourages both parties to build a relationship, there is a possibility that mentors find it difficult to give their mentees constructive feedback and ideas due to the fear of being misinterpreted as being too harsh or too patronizing of the younger mentees; and also mentees, due to the respect for the status of their mentors, might find it difficult to provide an unbiased evaluation of the feedback they get from their mentors. Our results suggest that having both the mentors and mentees interact in double-blinded relationship online would encourage them to give constructive and unbiased feedback to one another, which will in the long run increase the helpfulness and subsequently, the level of satisfaction that they both derive from the mentoring relationship.

3. Are incentives required for providing high quality feedback and if yes, what incentives?

When asked in the final questionnaire what motivated them to provide high quality reviews, 50% of the students stated that they were motivated by the feedback from the authors of the essays they reviewed (self-improvement; intrinsic); 33% - by the marks given by the marker (extrinsic), and the remaining 17% said they were motivated by just trying to be helpful to their colleagues (intrinsic, pure altruism). Also, when asked if they viewed the reviews provided by other reviewers of the same papers that they reviewed, the majority (71%) stated that they did, while 29% did not bother to check them. Some of them said that they did so out of curiosity and to know how they fared compared to other reviewers in their group. For example, one of the comments provided is:

- "Usually to just compare with my review out of curiosity." (Study participant)

This suggests that peers like to compare their reviews with other reviews of the same essay, which they review, in order to validate their understanding and criticism of the essay.

The mechanism contained incentives that could trigger extrinsic motivations for mentors (reviewers). Reviewers (peer mentors) received feedback on their reviews from the authors of the papers (the mentees). Also, they were able to see the reviews given by the other reviewers/mentors on the same papers that they reviewed and are allowed to give feedback on the usefulness of the other reviews. According to Bandura's self-efficacy theory, the feedback would enable them to learn and improve the quality of their reviews (Bandura, 1986; Grusec, 1992). In addition there was extrinsic feedback provided to reviewers, in terms of the grades of the reviews assigned by the maker (TA, instructor, super-mentor) as reported in tables 4.3 and 4.4.

4. What metrics can we use to measure the success of a mentoring relationship?

We did both implicit and explicit measures of the mentoring sessions. The objective of mentoring relationship in the class is to help participants learn to improve their writing and critique skills. Following the principle that one understands something best when one tries to help (in our case criticise) somebody else (Gartner *et al.*, 1971), we proposed group peer mentoring, i.e. every mentee is also a mentor. The goal was that after the session everyone would have benefitted and learned from the experience, in both roles, as a mentee and as a mentor. In our case study, students as authors were expected to learn how to write better, and as reviewers – how to criticize constructively and helpfully the work of others. We asked all participants (mentors and mentees) to provide the feedback on their learning experience, which constituted the explicit measure.

The final questionnaire contained questions probing whether the students felt their writing skills improved as a result of the peer-review sessions (i.e. that the session was useful for the mentees), and whether such a system should be used in other classes. 100% of the students said that the system should continue in the Ethics and IT class, while 83% of the students agreed that peer-review of course work should be used in other classes. Some of the suggestions provided by the students are:

- *"Peer review should be given a stronger emphasis in second and third year courses."*
(Study participant)
- *"Any classes with writing"* (Study participant)

When asked if, as reviewers they learned from seeing the other reviews, all of the students answered affirmatively. The students were asked specifically, if the peer review exercises have helped to improve their critique skills and 86% of them agreed. Some comments given by the students to this question are:

- *"Getting to compare different essays from different people gives you a larger amount of content to base your reviews on. Now I can more easily judge what is a poor essay and what is a good essay, because I have now read more poor, decent, and good essays than before."* (Study participant)
- *"I've become more immediately aware of problems in writing."* (Study participant)

To verify whether the actual actions of the students confirm their responses to this question, we analyzed the history of review versions and discussion comments from the EasyChair peer review system, and we found that at certain times, in 4 out of 7 review sessions (57.14%), some reviewers updated their reviews (edited their reviews) after submitting the first version of their review and seeing the other reviews of the paper.

The questionnaire also asked the mentees, if they think that the competence of their mentors (reviewers) improved over the course of the 7 review sessions. The majority (71%) answered "yes, 14% - "no" and 14% were not able to say if there was an improvement.

Our results showed that peers as mentors and mentees perceived that they learn from the mentorship experience they had with the modified peer review system. 71% of the students in the role of authors (mentees) confirmed over the many questionnaires during the term that their writing skills improved as a result of using the peer-review system and they all (100%) encouraged the continuous use of the system of course work in the Ethics & IT class and other classes that involve writing and programming. 86% of the students in the role of reviewers (mentors) confirmed that their critiquing skills improved. An analysis of the reviewers' activities also confirmed that some reviewers made changes to their reviews after seeing other reviews of

the same articles, which showed that they discovered points they missed in the other reviews. The main objective of introducing the group peer mentorship system, implemented as a modified peer-review system, was to help mentees improve their writing skill and mentors improve their critique skill. While the learning outcomes, measured by their grades did not show any significant pattern of change or improvement, the student questionnaires confirmed that their perception was that they learned and improved their skills, both as authors and as reviewers.

4.3.2 Case Study - Ethics and IT, 2014

We did another small scale field study with the same Ethics and IT class in the following year (winter 2014). Our goal was to know if our modified peer review system could support group peer mentorship to help students improve their writing and argumentation skills. This study lasted 6 weeks, during which students had to submit a weekly essay on the topic discussed in the class during the week. There were 5 short essays and a long essay submitted at the end of the term. In all, there were 6 peer review sessions, supported with the SWORD peer review management systems (Cho and Schunn, 2007). We decided to use SWORD for this experiment, because it implements a provision for back-evaluation of reviews by the authors, which is a crucial part of our modified peer review system. There were 10 participants, 8 of which were Canadians (6 male and 2 female) and 2 were international students. We did the initial grouping of students using the same approach that we used for the Ethics and IT class of 2013. The students were given an initial calibration task, which was a short essay and was graded to find out their writing skill levels. The students' grades from this calibration essay were used to group them and their group assignment changed every week based on their grades from the previous peer review session. However in this study, participants were encouraged to write their essay and review their peers' essay using pseudonyms. The SWORD peer review system had provision for grading peer reviews using natural language processing techniques. The weekly grade from the SWORD peer review system was also released to peers before they engaged in the next peer review session. At the end of every peer review session, students had to give a back-evaluation of the reviews they received, which probed their satisfaction with the quality and helpfulness of the reviews of their essays. In our final exit questionnaire, we asked the same questions to the students to gauge their perception of learning from the peer-review. The findings from this study are discussed below.

1. Does the diversity approach used to group peers achieve successful mentoring sessions?

To evaluate their overall perception of helpfulness of the peer review, participants as authors (mentees) provided the back-evaluation of each review that they received. At the end of each peer review session, participants in both roles provided their perception of the helpfulness and their learning from the peer review session; and they also had to fill an exit questionnaire at the end of the last (6th) peer review session, which probed their overall satisfaction with all the peer review sessions using certain criteria (see Table 4.5 & Appendix B). The mean and standard deviation of their responses are presented in Table 4.5. The results show relatively high average for the levels of agreement with the positive statements and the relatively low average for the levels of agreement with the negative statements, except for "too polite" that the respondents did not provide a response for. Also, there were relatively low standard deviation for the positive statements and relatively high standard deviation for the negative statements. This shows that the students tend to agree with the positive statements and were satisfied with the reviews that they received from their peers. This result also supports the hypothesis that the choice of diversity of group members is effective in achieving a successful mentoring session.

Table 4.5: Satisfaction levels of authors with the reviews they received according to the exit questionnaire (students' agreement with statements expressing the criteria is reported on a scale of 1 (lowest) to 10 (highest))

<i>Criteria</i>	<i>Mean</i>	<i>Standard deviation</i>
Competent	10	1.31
Constructive	10	1.42
Provided detailed suggestions	7.5	2.01
Provided useful correction to style and grammar	7.5	2.45
Too polite	2.5	1.98
Not too helpful	5.0	3.03
Not too thorough or detailed	5.0	3.69
Too negative	0.0	0.00
Lacked substance	2.5	3.85
Didn't always understand the point I was trying to make	2.5	3.56

2. Are incentives required for providing high quality feedback and if yes, what incentives?

We had a designated marker, whose identity was known to all the peers. Due to the workload on the marker, the marker only marked their weekly essay submission but did not mark their

reviews (unlike in the 2013 experiment). However, the SWORD peer review system graded the student's reviews, using NLP techniques (Xiong *et al.*, 2010) and they were able to see their grades from the SWORD system, in addition to the peers' feedback. In the exit questionnaire, we asked the students what motivated them to provide high quality reviews with options like the marks given by the SWORD peer review system, feedback from the authors of the essays that they reviewed and being helpful to their peers. 75% of the students said they were motivated by the feedback from the authors of the essays that they reviewed (intrinsic, self-improvement), while the remaining 25% believed that they were motivated by the marks given to their reviews by the SWORD peer review system (extrinsic). We probed further to know if they found these marks and the feedback from the author similar. 75% said that they found the marks and author's feedback similar; while 25% said they were not similar.

Table 4.6 below gives a summary of the marks given to the final essays by the marker. Although, peers seem to be satisfied with the reviews and the overall peer review approach, we discovered that their marks did not follow a particular trend but rather fluctuated over the weeks.

Table 4.6: Peers' grades from their final essay

<i>Users</i>	<i>Grade from Essay 1</i>	<i>Grade from Essay 2</i>	<i>Grade from Essay 3</i>	<i>Grade from Essay 4</i>	<i>Grade from Essay 5</i>	<i>Grade from Essay 6</i>
A	85.00	100.00	95.00	93.00	100.00	80
B	100.00	90.00	100.00	85.00	75.00	50
C	98.00	95.00	93.00	90.00	65.00	85
D	90.00	95.00	90.00	100.00	85.00	95
E	80.00	75.00	90.00	50.00	85.00	55
F	80.00	65.00	80.00	40.00	95.00	87
G	80.00	70.00	70.00	82.00	50.00	60
H	100.00	80.00	90.00	68.00	90.00	86
I	-	75.00	95.00	90.00	-	80
J	70.00	75.00	75.00	78.00	60.00	45
Average grade	87.00	82.00	87.80	77.60	78.33	72.30

We also asked if they checked the reviews, done by the other students, on the same essays that they reviewed. 75% said that they didn't check the reviews, while 25% said they did. We believe this can be attributed to the design of the SWORD system which was used for the experiment. One of the comments provided was:

- *"I didn't realize I could until I saw this question"* (Study participant)

However, we could not check if those who said that they checked the other reviews made changes in their own review, because the SWORD peer review system over-writes every submission with the new version and does not provide the history of changes made.

3. What metrics can we use to measure the success of a mentoring relationship?

We did an explicit measure of success of the peer review or mentoring session. Since the objective of the group peer mentorship sessions is to help students improve their writing and critique skills, so we decided to ask peers in both roles as mentors and mentees (of which every peer is both in our experiment) to provide feedback on their learning experience (which is the explicit measure). 75% believed that there has been some improvement in their writing skill as a result of the peer review assignments; while 50% said that their critique skill as a reviewer has also improved. One of the comments given by the participants is:

- *"Was a really neat way to write essays and improve our writing skill"* (Study participant)

However, they all agreed (100%) that the system of coursework (peer review) should continue and they gave suggestions as:

- *"In any essay-based class"*(Study participant)
- *"I think this might work well in one of the more theory heavy classes..."* (Study participant)

4.3.3 Comparison of findings from the two experiments (Ethics 2013 and Ethics 2014)

Here, we present the commonalities and differences in our findings from the Ethics 2013 and Ethics 2014 experiments. In the Ethics 2013 experiment, participants found the feedback from their peers, as the authors of the papers that they reviewed, to be more motivating than the marks given on their reviews by the marker. In the Ethics 2014 experiment, the marker did not grade their reviews. However, their reviews were graded automatically by the SWORD system and the participants also found the feedback from their peers to be more motivating than the feedback from the system.

In the Ethics 2013 experiment, we tested for anonymity and its effects on the quality of essays and reviews produced by the participants over time. We found out that anonymity encouraged peers to be critical in their reviews, which translated to increasing the quality and the helpfulness of their reviews. We did not test for anonymity in the Ethics 2014 experiment. In both experiments, we found out that the participants were satisfied with the modified peer review system by asking for feedback from peers in both roles - as authors and reviewers. However, their satisfaction did not translate to improved performance, because the marks given by the markers in both cases did not show a sustained improvement in their writing skill.

4.3.4 Case Study - Research Methods (CMPT 880/890) Class, 2013-2014

We ran another study in a graduate class on Research Methods (CMPT 880/890), beginning from fall 2013 to the end of winter term 2014. One of the goals of the class is to help students build research, writing and review skills, which are important for every emerging or early researcher. Therefore, it is a writing-intensive class and our goal of using the class for the experiment was to find out if our modified peer review system can support group peer mentorship to aid the writing and review skills of the students. There were four writing assignments submitted as coursework throughout the two-term class focusing on summarising, or critiquing a research paper in the area in which each student was doing research. In addition, at the end of the class, the students submitted individual final papers presenting their proposed research projects, which were significantly longer than the four previous writing assignments. The students also used the SWORD peer review system to manage the submission and the peer review of their assignments. The SWORD peer review system had provision for grading peer reviews using natural language processing techniques. The weekly grade from the SWORD peer review system was released to peers before they engaged in the next peer review session. We gave an exit questionnaire at the end of the experiment (See Appendix B), to gauge their perception of learning from the peer review process and to answer our research questions (see section 4.1). The participants in this study were 24 graduate students; 4 Canadians and 20 international students, mostly of Asian origin. The results from this study are reported below.

1. Does the diversity approach used to group peers achieve successful mentoring sessions?

Due to the fact that the participants were graduate students working in different research areas, we surmised that reviewing the assignments of peer students from different research labs would

be challenging for them. That is why we modified the review assignment procedure slightly, but still preserving the diversity principle. The students were initially clustered in groups according to their research group affiliations, and sub-groups were created from these clusters using the initial calibration essay. At the end of each peer review session, participants were asked to give a back-evaluation of the reviews they got on their assignments, through the SWoRD system. See Table 4.7 for the means and standard deviation of their responses. As seen from the table, there are moderately high average values for the positive statements, while there are relatively low average values for the negative statements. Also, there are relatively low standard deviation for the positive statements and relatively high standard deviation for the negative statements. Also, we noticed zero means and standard deviations for "too negative" and "lacked substance". Yet the positive means are lower than those reported in the last two Ethics Class case studies. We see particularly low values for the perceived competence of the reviews and the usefulness of grammar and style corrections. The lower perceived competence may be due to the fact that the reviewers had usually not read the paper on which the author was writing their assignment, as each student wrote their assignment on an individually chosen paper. The lower perceived usefulness of grammar and style corrections was due to the fact that the overwhelming majority of the class consisted of non-native English speakers.

Table 4.7: Satisfaction levels of authors with the reviews they received (agreement with statements expressing the criteria, using a scale of 1 (lowest) to 10 (highest))

<i>Criteria</i>	<i>Mean</i>	<i>Standard deviation</i>
Competent	5.0	1.05
Constructive	8.0	1.67
Provided detailed suggestions	5.0	1.10
Provided useful correction to style and grammar	4.0	1.15
Too polite	1.0	2.96
Not too helpful	1.0	2.99
Not too thorough or detailed	1.0	3.10
Too negative	0.0	0.0
Lacked substance	0.0	0.0
Didn't always understand the point I was trying to make	2.0	3.26

2. Does anonymity enhance the quality of mentorship?

This experiment started out with participants using their pseudonyms. However, the anonymity of the student identities was broken at the third review session, when participants were able to identify each other. We asked the participants to rate their overall satisfaction with their weekly

reviews (see Figure 4.2). Their responses showed a 69% average satisfaction with their reviews. There was also a drop in their satisfaction levels from 75% to 60% in the third review session (RS3), when the anonymity was broken. We can attribute this drop to both the revelation of their identities at this time and the increased workload from their course work and other courses they might be taking, towards the end of the fall (1st) term, because the experiment in this class extended from the fall (1st) to the winter (2nd) term. On the overall, 70% of the participants believed that the quality of their reviews improved over time.

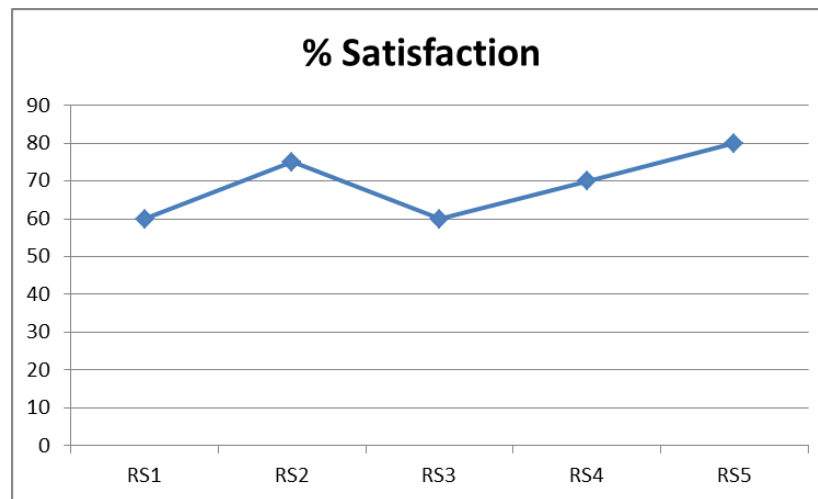


Figure 4.2: Authors' satisfaction with the reviews received in each reviewing session
(Research methods, 2014)

3. Are incentives required for providing high quality feedback and if yes, what incentives?

We asked the participants, in the final questionnaire, what motivated them to provide high quality reviews, with options like the marks given by the SWoRD peer review system, feedback from the authors of the essays that they reviewed and being helpful to their peers. 20% said they were motivated by the feedbacks from the authors of the papers which they reviewed, while 80% said they were just being helpful to their colleagues. This shows a huge deviation from our results in the last two experiments. We attribute the difference between this result and the results from the last two experiments to the difference in the background and general goals of the participants. Participants in this experiment were mostly international students from countries with collectivist culture (Hofstede, 2010); while the participants in the other two experiments were mostly Canadian students, coming from individualist culture. However, this claim is still open to further experiment. When asked if they checked the other reviews of the same paper that

they reviewed, 50% said they did, while 50% said they did not. One of the comments by the 50% that checked the other reviews is:

- *"Reading reviews by others was probably the most helpful thing"* (Study participant)

We attribute this result to the design of the SWoRD peer review system, which they used for the experiment. One of the comments by the 50% that did not check the other reviews is:

- *"Not sure I knew I could do that"* (Study participant)

4. What metrics can we use to measure the success of a mentoring relationship?

We did an explicit measure of success for the peer review experiment by asking each participant to provide feedback on their learning experience. They all (100%) said that there has been some improvement in their writing skill as a result of the peer review coursework and 70% believed that their critique skill as a reviewer have improved. Some of the comments that they gave are:

- *"I think I have specifically learnt to think about the reader more, and to present things in a way that is accessible to them"* (Study participant)
- *"Just having to do so much writing is helpful. Reading other papers was good too. Some reviews I got from others really opened my eyes to common things I was doing wrong, and I think I fixed them in the end"* (Study participant)

70% of the participants said that the system of course work should be continued and extended to other classes too.

4.3.5 Case Study – Ethics & IT Class, 2016

We did a case study with an ethics and IT class from January to April 2016. The goal of this study was to help the participants in group peer mentorship improve their writing and review skills. The study lasted 10 weeks, and there were 17 participants in the study. Out of the 17 participants, four were female and 13 male. Our participants were split across different ethnic backgrounds. We had eight Caucasians, five Asians, three Africans, and one Latin American. We had a designated marker who graded their essays and final paper. There were seven writing assignments; six were peer-reviewed while one assignment (the fourth assignment) was not peer-reviewed but was only graded by the marker. The marker's feedback was both formative and

summative because it contained comments suggesting areas of improvements for the authors and a grade that was part of their final class grade. In this study, we gave out weekly and exit questionnaires to our participants (Appendix C, parts 3 and 4), but we did not have sufficient data to answer the first two questions as we did in the last three studies. However in this study, we were able to measure the success of the mentorship and ask a further question to confirm if the introduction of back-evaluation has actually helped our modified peer review process to fulfill its formative objective. The findings from this study are discussed below.

1. What metrics can we use to measure the success of a mentoring relationship?

We did an implicit measure of the success of the mentoring relationship in this study by looking at the changes in the writing and review skills of our participants. We had a designated marker, whose identity was known to all the peers. The marker marked their weekly essays and their final project paper. Because of the high workload of the students who participated in this study, we were unable to receive a substantial response from the weekly questionnaire. However, participants provided both summative and formative feedback on their peers' essays and provided the back-evaluation of their weekly reviews through the Peerceptiv system. For every peer review session, participants received feedback from their peers and the designated marker except in the fourth assignment when the assignment was only marked by the marker. Table 4.8 below shows the progression of participants' grades over the seven assignments.

Table 4.8: Progression of participants' grades as judged by the marker

<i>Participants</i>	<i>Week₁</i>	<i>Week₂</i>	<i>Week₃</i>	<i>Week₄</i>	<i>Week₅</i>	<i>Week₆</i>	<i>Week₇</i>
A	9	8.5	8.5	9	9	9.5	8
B	8.5	10	10	10	10	10	10
C	6	8	8.5	10	8.5	8	6.5
D	6.5	8.5	7	7.5	9	7.5	7.5
E	8	9.5	10	10	Null	Null	Null
F	8	7	8.5	9	9.5	9	9
G	0	Null	8	6.5	7.5	9	9.5
H	9	9	8.5	7.5	7.5	9.5	9.5
I	5	7	8.5	7	7.5	9	9
J	9	9	10	9	10	9.5	10
K	8	7.5	9	10	8	8	7.5
L	7.5	9.5	10	8.5	8	9	10
M	8	7	Null	6	7	3	7
N	9	9	9	7.5	8.5	8	8

<i>Participants</i>	<i>Week₁</i>	<i>Week₂</i>	<i>Week₃</i>	<i>Week₄</i>	<i>Week₅</i>	<i>Week₆</i>	<i>Week₇</i>
O	0	4	10	Null	Null	9.5	7
P	8.5	8	9	9	9	9.5	10
Average marks	7.33	8.1	8.97	8.43	8.5	8.53	8.57

Although their individual marks varied across the seven assignments, there was an increase in the average marks of all participants from the first peer review assignment (7.33) to the third peer review assignment (8.97). However, their average marks dropped from 8.97 to 8.43 in the fourth assignment when they did not peer review each other’s essays. We also observed an increase in their average marks from the fifth assignment, when they resumed the peer review exercise, at a steady rate until the last peer review assignment. At the end of the seventh peer review assignment, participants submitted a term paper, which is a longer version of their weekly essays, and peer review was not mandatory for this paper. However, 75% of our participants chose to submit their papers for peer review. One of the study participants provided the comment below to solicit a submission waiver when the student could not meet the initial deadline for submission.

“I highly value the feedback that we have received through peer evaluation thus far in the class”
(Study participant).

We also calculated the average of the overall evaluation (ratings) that peer reviewers gave the authors for each weekly assignment, except for the 4th assignment, which was not peer-reviewed (see Table 4.9). There is also no sustained trend from the first to the seventh assignment.

Table 4.9: Average ratings of the weekly essays by the peer reviewers

<i>Week 1</i>	<i>Week 2</i>	<i>Week 3</i>	<i>Week 5</i>	<i>Week 6</i>	<i>Term paper</i>
5.68	5.08	6.01	5.5	5.74	5.59

We asked participants to provide the back-evaluation of their reviews by rating the helpfulness of their reviews on a Likert scale of 1 to 7 for the six peer-reviewed assignments. Table 4.10 shows the average helpfulness values of the peer reviews for the six peer-reviewed assignments, as judged by the authors of the essays that were reviewed. We observed that the average helpfulness values dropped from 4.14 in the first assignment to 4.11 in the second

assignment. However, we observed a weekly increase in the average helpfulness values from 4.11 in the second peer review assignment to 4.13 in the third assignment and finally to 4.54 in the sixth peer review assignment.

Table 4.10: Average helpfulness of the weekly peer reviews

	<i>Review 1</i>	<i>Review 2</i>	<i>Review 3</i>	<i>Review 4</i>	<i>Review 5</i>	<i>Review 6</i>
Average Helpfulness	4.14	4.11	4.13	4.22	4.29	4.54

The goal of this study was to support peers in group peer mentorship to improve their writing and review skills. The results from this study show that there was no particular trend in the judgement of participants' writing skill as authors, as shown in Table 4.8, and also it shows an improvement in the helpfulness of their reviews as peer reviewers, as shown in Table 4.10. A likely explanation for the lack of marked improvement in the essay grades in the last essay is the very high load on the students (in fact the student complained in their course evaluations that the load of 7 writing assignments over the term was extraordinarily high).

2. Does the back-evaluation encourage reciprocation of ratings?

In this case study, we asked an additional question related to the fairness of ratings. Authors gave back-evaluation of the reviews they received. This represents a helpfulness rating of the reviews as perceived by the authors, who are the recipients of the reviews. To mask their identities for double blind review and also to be able to accrue their peer ratings over different peer review sessions, participants, both as authors and reviewers, were encouraged to use pseudonyms throughout the six peer review sessions. We expected that they would be willing to give truthful and helpful feedback to their peers without feeling compelled to reciprocate either good or bad ratings from their peers. To measure this, we calculated the correlation between the review ratings that the peer reviewers gave to the authors and the back-evaluation ratings that the authors gave their peer reviewers for six assignments. We omitted the fourth assignment because there was no peer review for that week. We expected that the correlation values would be strongly positive if there were a reciprocation of ratings (see Table 4.11).

Table 4.11: Correlation between weekly review ratings and back-evaluation ratings - Ethics and IT, 2016

	<i>Correlation</i>	<i>R-Squared (R²)</i>
Assignment ₁	-0.2218	0.0492
Assignment ₂	0.2573	0.0662
Assignment ₃	0.6226	0.3876
Assignment ₅	0.0865	0.0075
Assignment ₆	0.0974	0.0095
Assignment ₇	0.2921	0.0853

In the six peer review sessions, we observed a mix of positive, negative, weak, and strong correlations. Therefore, we calculated the R-squared to determine the strength of the relationship between the review and back-evaluation ratings. Our results showed that the review rating is a weak predictor of the back-evaluation rating. For example in Assignment₃, where the R-squared value is the highest, 0.3876, this means that 38.76% of the variance in the back-evaluation ratings can be explained by the peer review ratings.

4.3.6 Case Study – Masters of Professional Accounting Class, 2015

We conducted a study with professional accountants, who were also registered students of the Masters of Professional Accounting Program at the Edwards School of Business at the University of Saskatchewan. Thirty-seven (37) students consented to take part in the study. However, only seventeen (17) of them completed the study. This study differs from our previous studies because it comprised an initial survey to measure the interpersonal trust score of each participant, using Rotter’s interpersonal trust / distrust scale (see Appendix C, part 2), and two peer review sessions using an existing online peer assessment system, called Peerceptiv (Peerceptiv, 2016), during which the participants analyzed two audit cases and also reviewed each other’s analyses. The goal of this study was to support participants in group peer mentorship to improve their audit, analysis and peer review skills. We found from the literature that peers’ trust in each other’s ability to perform could influence the success of the group peer mentorship (Hodges, 2009). However, the existing literatures lack an explanation of the dynamics of trust in the mentoring relationship. Therefore, with the inclusion of the measure of peers’ interpersonal trust score, we hope to find the correlation between their interpersonal trust scores and their rating tendencies as both mentor and mentee, in addition to answering the research questions that we set out in section 4.1.

Participants were assigned into groups of at most four peers for each session. The first assignment was based on their previous course average, instead of an initial calibration task, because due to their working full time, they did not have time to complete three tasks. In the second peer review session, participants were re-grouped using their peer ratings from the first peer review session. They used the Peerceptiv peer review system to manage submission and the peer review of their audit analyses. At the end of each of the two peer review sessions, participants were asked to evaluate their experiences both as authors (mentees) and reviewers (mentors), using similar questionnaires to the ones that were used for the previous four experiments (see Appendix C parts 3 and 4). We also gave an exit questionnaire at the end of the experiment, to gauge their perception of learning from the peer review process and to answer our research questions (see section 4.1). However, it was difficult to get tangible survey responses that could answer our research questions (section 4.1) from the participants due to their work schedule. Therefore, we could not answer the same research questions as we did in the last four experiments. However, we found some interesting trends between their interpersonal trust scores and the peer review ratings that the mentors gave their mentees, and also between their interpersonal trust scores and the back-evaluation ratings that mentees gave their mentors, from the log data in the peer review system. With these, we hope to answer the following research question:

“How do peers’ interpersonal trust scores affect their ratings behavior both as mentors and as mentees?”

To frame our analyses, we proposed two null hypotheses:

H₀₁: *Mentees with high trust scores will be generous in their ratings and offer high ratings in the back-evaluation of their mentors.*

H₀₂: *Mentors with high trust scores will be generous in their ratings and offer high ratings in their review feedback.*

To address these hypotheses, we asked participants to complete, in addition to their consent form, an initial survey by Julian Rotter (Interpersonal trust scale, 2015) to measure their interpersonal trust score (see Fig. 4.3). We found that the average trust scores of the 17 participants that completed the study is 75.06, which Rotter defined as a mild level of interpersonal trust score, and the standard deviation is 6.79. This trust score is consistent with

other research involving financial statement auditors, where the participants were also found to have, on the average, a mild level of interpersonal trust score (Quadackers *et al.*, 2014).



Figure 4.3: Participants' interpersonal trust score from Rotter's scale

For both peer review sessions, the peer mentors provided feedback based on three criteria:

- 1) consideration of engagement issues and risks (*Qn1*),
- 2) evaluation of accounting choices (*Qn2*), and
- 3) consideration of materiality, audit approach and audit procedures (*Qn3*).

We collected both the individual and average peer review ratings that participants as mentors gave their mentees and the back-evaluation ratings that the participants as mentees gave their mentors, all in a series. These values were each compared with their interpersonal trust scores.

1. **H₀₁**: *Mentees with high trust scores will be generous in their ratings and give high ratings in the back-evaluation of their peer mentors.*

We calculated the correlation between mentees' (authors') trust scores and the back-evaluation ratings that they gave their peer mentors (reviewers) (see Table 4.12). Our results showed that there is a moderate negative correlation (-0.2047) between their trust scores and the back-evaluation ratings that they gave their peer mentors. That is, the higher their tendency to trust, the lower the ratings they give their peer mentors. Although, this result looks contrary to the definition of trust by Mayer, Davis and Schoorman (1995) that a trustor would be willing to be

vulnerable and expect to take risk due to the uncertainty and inability to control the other party; it further reinforces the findings of (Rotter, 1980; Yamagishi and Kikuchi, 1999) that trustworthiness is not synonymous with gullibility. That is, high trustors tend to have high expectation of other people. In our case, the more trusting mentees have high expectations of the performances from their peer mentors, therefore, they penalized their mentors more when their expectations of their feedback was not met. It further explains why, despite the fact that our participants have on the average a mild level of interpersonal trust (75.06), the average rating that they gave their peers for their reviews was 1.6 out of the total possible rating of 7. We also split the sample into quartiles based on participants' trust scores. Table 4.13 shows the mean and standard deviation of the review and back-evaluation ratings for each quartile of the sample.

Table 4.12: Correlation: trust scores vs. back-evaluation & trust scores vs. average review ratings

	<i>Trust score vs. Back-evaluation</i>	<i>Trust score vs. average reviews ratings</i>
Correlation	-0.2047	0.0208
p-value	0.22	0.47

Table 4.13: Mean and Standard deviation of trust scores and ratings for each quartile of trust scores

<i>Quartiles</i>	<i>Trust Scores</i>		<i>Review Ratings</i>		<i>Back-evaluation</i>	
	<i>Mean</i>	<i>St. Dev</i>	<i>Mean</i>	<i>St. Dev</i>	<i>Mean</i>	<i>St. Dev</i>
Q1(72)	70.54	1.82	4.00	2.15	2.08	2.16
Q2(76)	73.00	3.35	5.15	1.17	2.23	1.89
Q3(78)	76.92	0.73	5.23	1.12	1.08	1.54
Q4(92)	84.15	6.02	4.92	1.44	1.31	1.94

These results do not support the hypothesis that mentees with high interpersonal trust scores will be generous and give high ratings in the back-evaluation of their peer mentors.

2. **H₀₂:** *Mentors with high trust scores will be generous in their ratings and offer high ratings in the review feedback that they give their mentees.*

We calculated the correlation between the average reviews feedback and the reviewers' trust scores. Our result showed that there is a very weak positive correlation between their

interpersonal trust scores and average review ratings (see Table 4.12). That is, a highly trusting reviewer (mentor) will most likely offer high ratings to their author (mentee). We see this as a benevolent act, which Mayer *et al.* (1995) described as one of the factors that contribute to the development of trust in a trustee. However, the correlation is weak, which is an indication that more trusting people tend to give slightly positive ratings as mentors to the work of their mentees. We also calculated the correlation between their trust scores and the individual review ratings for each criterion, *Qn1*, *Qn2* and *Qn3* that was used in the peer review (Table 4.14). The results show that *Qn1* (0.0344) and *Qn2* (0.0186) have weak positive correlation with their trust scores. *Qn3* and trust score show a very weak negative correlation (-0.0023). We also computed the R-squared values, which showed that for each criterion used in the peer review, the trust score was a weak predictor of the ratings that they gave their peers as mentors (see Table 4.14).

Table 4.14: Correlation: trust scores vs. individual review ratings

	<i>Trust score vs. Qn1</i>	<i>Trust score vs. Qn2</i>	<i>Trust score vs. Qn3</i>
Correlation	0.0344	0.0186	-0.0023
R-Squared(R^2)	0.0012	0.0003	5.194×10^{-6}

These findings, therefore, do not support the hypothesis that mentors with high trust scores will always be generous in their ratings and therefore, offer high ratings to their mentees.

4.4 Conclusion

We found out that the existing online mentorship systems still pair mentors and mentees manually in one-to-one matching, which has been proven not to be cost and time effective, and its efficiency could be affected by the personal attributes of the mentors and the mentees involved. We proposed to use peer-review model to implement online group peer mentorship as alternative to existing one-on-one online mentorship programs. Analysis of the stages in the traditional peer review approach showed that they were not sufficient to fulfill the requirements of our proposed peer group mentorship system, and we proposed a modified double-blind peer-review model that differs from the usual one in two aspects: peer-assignment in a group of peers with different expertise levels, and including additional steps in the peer-review process where mentees (authors) evaluate the competence and helpfulness of the reviews they receive from the mentors, and the peers in both roles evaluating their experience with the peer review process.

The findings from the first four case studies suggest that the proposed mechanism for online group peer mentorship, based on online peer-review, can be usefully applied in a class to help students improve their writing and critiquing (reviewing) skills. Our findings show that mentees found the feedback from their peers more motivating and helpful than the feedback they got from a senior person (the marker/TA/Instructor) or the SWORD system. From Tables 4.2, 4.5 and 4.7, it is clear that the students appreciated the feedback they received on average and Figures 4.1a-b and 4.2 show that students were generally satisfied with the reviews received each week.

Specifically, our findings confirmed the advantages of anonymous review as it allows peers in a mentor role to give more thorough, critical and honest/unbiased reviews. Our results also suggest that providing an opportunity for mentees to evaluate and comment on their mentor's feedback (review) can help the mentors to improve their skills for constructive criticism and also motivates intrinsically the mentors to give better reviews. This result is in agreement with Albert Bandura's self-efficacy theory, which emphasizes people's perception of their ability to change, as a motivating factor for self-improvement. In our case this perception is facilitated by the feedback from their peers. Our findings also relate to Robert Sears' social learning theory that people are influenced by rewards, which serve as punishment or reinforcement of their actions. In our case, the feedback given by mentees to the mentors constitutes their rewards, which we found to have helped them to improve their review and critic skills. Our findings also suggest that reviewers (mentors) learn from being able to see the other reviews (feedback) to the same mentees or papers. The activities of peers as mentors should be made visible to one another, as it would provide opportunity for peers to compare the feedback they give on the same question, become exposed to different viewpoints and learn how to be better mentors.

In a typical online peer review system, success is based on the acceptance rate of papers (typically the lower the acceptance rate, the higher the quality of the conference). This is one-sided, as it only measures success from reviewers' and editors' points of view. We believe that the success of a peer review or mentoring session should be determined by the helpfulness of feedback provided to authors (mentees) by reviewers (mentors), and the independent data from the peer review system. The success of a mentorship system should be based on feedback from all participants (both mentors and mentees) about how much they learned from the experience. Evaluation from mentees or authors of the quality of feedback provided by the mentors or

reviewers would be essential in a pseudonym-based peer-review system, as it can help mentors or reviewers to build reputation, and would both serve as extrinsic motivation for doing a good job.

From the last two experiments described in this chapter, we were able to explore some factors that could affect peer ratings in group peer mentorship – reciprocation of ratings by peers, and peers interpersonal trust scores. We found that, by including back-evaluation of review, peers were not trying to reciprocate their ratings but were just trying to provide honest opinions of their reviews. We found also that peers' rating behavior is influenced by their trust score and it depends on the roles that they assume in the mentoring relationship. That is, the higher the tendency of mentees to trust, the lower the ratings they give their peer mentors, and highly trusting mentors will most likely offer high ratings to their mentees.

CHAPTER 5

LARGE SCALE CASE STUDY WITH TEACHERS IN CHILE

In July 2013, we presented our idea of supporting group peer mentorship with a modified peer review system at the conference on Artificial Intelligence in Education (AIED) (Adewoyin and Vassileva, 2013). During our interaction with other researchers in the field, we were presented with the opportunity to conduct a large scale experiment with teachers in Chile, using our modified peer review system. In this chapter, we present the results of the large scale study with the teachers in Chile.

5.1 Introduction

In 2013, the Chilean government funded CIAE, the Educational Research Center of the University of Chile, to conduct day-long seminars to help school teachers from the entire country to implement the new national curricula in three disciplines – Mathematics, Music and Language. In these seminars, international and national speakers of each discipline gave lectures on the new didactic methodologies. About half of the four thousand teachers that attended the seminar attended in person, while the rest by video streaming. Then, 168 Mathematics teachers, 73 Music teachers and 43 Language teachers voluntarily engaged in the peer review study, which lasted two months. Participants in each subject group were divided into control and experiment sub-groups. We provided the modified peer review process (see Fig. 3.1 in chapter 3), which was implemented by the CIAE team in Chile. Participants in both control and experimental groups engaged in two rounds of peer review process. The teachers as peers provided all the ratings used in the data analysis. The experiment took place in Chile and due to a mix-up the collation of data for the Music and Language groups, we could only get the results of the controlled experiment for the Mathematics group. Therefore, we present a controlled study of using our modified peer review process to support the Mathematics teachers in section 5.3. In section 5.4, we present combined results for the three groups. The goal of the study was to find out if the modified peer review system could support the participants' understanding of the new didactic methodologies to implement the new curricula in three subject areas – Mathematics, Music and language. In order for peer review to support our mentorship model, we will like to explore the following questions.

1. Does the diversity approach used to group peers achieve successful mentoring sessions?

In this study, peers were initially grouped randomly for the first peer review session, and peers in the control group were again grouped randomly for the second peer review session. Although in the previous experiments (chapter 4), peers were given an initial calibration essay, which was used to judge their competence, we could only use the feedback from the first of the two peer review sessions in this study to judge their competence because we did not have a designated marker to mark the initial essay of the peers in this experiment. So, the feedback from their first essay was used to judge their competence and to group peers in the experimental group for the second peer review process. With this approach on the experimental group, we want to ensure that the high competent peers are grouped with peers that are weak; with the hope that they would be mentored and that the high competent peers would also learn from the experience.

2. Are incentives required for providing high quality feedback and if yes, what incentives?

We proposed that participants in the role of reviewers should get feedback from the authors on the reviews they gave them and also see the reviews given by other reviewers on the same paper. This, we believe, would allow them to learn from others and would also motivate them to improve on the quality of the reviews that they give. We also proposed that participants used pseudonyms for the mentorship in order to allow them to give critical and thorough feedback, and to also enable them accrue their feedback in form of reputation over time.

3. What metrics can we use to measure the success of a mentoring relationship?

In order to judge the overall success of the review session, we proposed that all the participants, both as authors and reviewers, provide the evaluation of the helpfulness and learning from the peer review process. Also, we proposed to ask if participants would like to recommend the use of the system in their schools.

5.2 Research Method

The participants were engaged in two peer review sessions. In the first peer review session, participants were randomly matched to review each other's works. In the second peer review session, participants were classified into two groups – control and experimental groups. Participants in the control group were randomly matched for the second peer review session. In

the experimental group, the matching was done using the diversity principle that we used in the previous experiments described in Chapter 4. The competence of the participants was evaluated from the peer-evaluation ratings results of the first peer review session. The goal was to measure the impact of using our approach of modified peer review, using the diversity matching approach, in improving their writing (in this case, drafting syllabus) and critiquing skills and also on their perception of helpfulness and learning from the system, by comparing the results from the experimental group with that of the control group at the end of the experiment. Out of the 168 Mathematics teachers that participated in the first peer review session, 74 teachers completed the experiment. Among the 74 Mathematics teachers, there were 40 participants in the control group and 34 participants in the experimental group (see Figure 5.1).

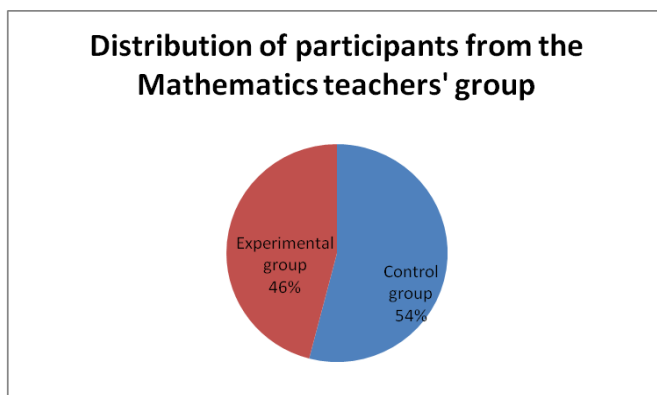


Figure 5.1: Distribution of participants from the Mathematics teachers' group

In both groups, their papers were evaluated through anonymous peer review process, with each participant in the position of both an author and a reviewer. Participants were encouraged to give constructive feedback and also made to rate the papers that they reviewed. In return, the author of each paper would also rate the helpfulness of the reviews of their paper, based on certain criteria included on the questionnaire used for the feedback (see Tables 5.4 and 5.5). Also reviewers (mentors) were asked to provide feedback on the reviewing session of each individual paper, whether they learned and improved their skills for constructive critiquing, from seeing the other reviews of the paper and from the ensuing discussion.

5.3 Case Study Results and Discussion – Mathematics Group

The experiment was conducted in Spanish. So, we had to use a translator to get the responses in English language. There were two peer review sessions, supported by their custom-made peer review system. Participants were asked to fill a short survey at the end of each session. There was no designated marker in this experiment. So the participants were asked to rate the papers that they reviewed and also to provide some summative feedback, and the participants as authors were asked to provide feedback on the quality and helpfulness of the reviews of their paper. The feedback was used in the second iteration to group the participants in the experimental group. With this study, we hope to find answers to the aforementioned questions (section 5.1) and come to a conclusion regarding the potential of our approach to support group peer mentorship, using the data from the teachers in the Mathematics group.

1. Does the diversity approach used to group peers achieve successful mentoring sessions?

In the experimental group, participants were matched for the second peer review session using the rating of their papers and their reviews from the first review session, while the participants in the control group were randomly matched for both peer review sessions. To determine the effectiveness of our group matching approach to achieve successful mentoring session, we asked participants as reviewers to rate the papers that they reviewed, from 1 -worst to 7 - best, using three questions. The mean of their responses in both the first and second peer review sessions are presented in tables 5.1 and 5.2. The peer ratings of the papers from the first peer review session did not show any advantage of the experimental group over the control group. In fact, the participants in the control group had a higher average rating (5.43) than the participants in the experimental group (5.33) and both groups were subjected to the same reviewer assignment condition in this iteration, which is random matching (see Table 5.1). However, in the second peer review session when participants in the experimental group were matched using the diversity approach, there was a noticeable improvement in the average of the peer rating of the papers of the participants in the experimental group (5.63) over the average of the peer ratings of the papers of the participants in the control group (5.33) (see Table 5.2).

Table 5.1: Average peer ratings of the participants' papers from the first peer review

<i>Question</i>	<i>Control Group</i>	<i>Experimental Group</i>
The speaker presented an idea that the participant intends to put into practice. How close is the participant' proposal to what was presented by the speaker?	5.8	5.7
Is it possible to put into practice what is proposed by the participant?	5.4	5.4
Does the participant introduce any new variation that would be relevant for them and other instructors?	5.1	4.9
Average	5.43	5.33

Table 5.2: Review feedback from the second peer review session

<i>Question</i>	<i>Control Group</i>	<i>Experimental Group</i>
The speaker presented an idea that the participant intends to put into practice. How close is the participant' proposal to what was presented by the speaker?	5.6	6.1
Is it possible to put into practice what is proposed by the participant?	5.3	5.6
Does the participant introduce any new variation that would be relevant for them and other instructors?	5.1	5.2
Average	5.33	5.63

To test the significance of our results, we conducted a paired-samples t-test on the data from the second peer review session (Table 5.3).

Table 5.3: Paired-Sample T-test on the Data from the second peer review session

		<i>Control group</i>	<i>Experimental Group</i>
Mean		5.33	5.63
SD		0.25	0.45
Mean difference	0.30		
p-value for paired t-test	0.12		
t-value	2.60		
Df	2		

Since our $p\text{-value} > 0.05$, paired-samples t-test showed that experimental group have no statistically significant different scores ($M = 5.63$, $SD = 0.45$) compared to control group ($M = 5.33$, $SD = 0.25$), $t(2) = 2.60$, $p = 0.12$.

We also asked the participants in both groups to back-evaluate their satisfaction with the reviews they received from their peers (from 1 - worst, to 7 - best) using certain criteria. The mean of their responses are presented in the tables 5.4 and 5.5.

Table 5.4: Back-evaluation of the reviews from the first peer review session

<i>Question</i>	<i>Control Group</i>	<i>Experimental Group</i>
Was the feedback pertinent?	5.5	5.6
Was the feedback constructive?	5.2	5.6
Did the feedback provide detailed suggestion of how to improve your paper?	4.6	5
Did the feedback provide reasonable suggestion?	4.8	5.2
Was the feedback polite?	4.8	4.9
Was the feedback provided in a friendly tone?	5.8	5.8
Evaluate the clarity of the feedback	5.2	5.5
Did the feedback provide interesting issues?	5.0	5.2
The corrections provided by the feedback are useful	5.0	5.3
Evaluate the overall comment provided by the feedback	5.3	5.4
AVERAGE	5.12	5.33

Table 5.5: Back-evaluation of the reviews from the second peer review session

<i>Question</i>	<i>Control Group</i>	<i>Experimental Group</i>
Was the feedback pertinent?	5.5	5.9
Was the feedback constructive?	5.3	5.9
Did the feedback provide detailed suggestion of how to improve your paper?	4.8	5.2
Did the feedback provide reasonable suggestion?	5.3	5.6
Was the feedback polite?	5.1	5.4
Was the feedback provided in a friendly tone?	5.9	6.1
Evaluate the clarity of the feedback	5.5	5.7
Did the feedback provide interesting issues?	5.1	5.2
The corrections provided by the feedback are useful	5.1	5.5
Evaluate the comment provided by the feedback	5.3	5.7
AVERAGE	5.29	5.62

As seen in tables 5.4 and 5.5, there is no change in the ratings of the overall comments provided by the reviewers in the control group from the first peer review (5.3) to the second peer review session (5.3) (see the 11th row of tables 5.4 and 5.5). However, there is a positive increase in the rating of the overall comments provided by the reviewers in the experimental group (from 5.4 to 5.7) (see the 11th row of tables 5.4 and 5.5). Although there is a difference in the average feedback of the control (5.12) and experimental (5.33) groups right from the first peer review session, this difference widens in the second peer review session (control group:5.29; experimental group:5.62) (see the 12th row of tables 5.4 and 5.5).

To see if there is a significant difference between the aggregate scores for all questions in the second peer review session, we calculated the paired-samples t-test (see table 5.6). The results of the paired-samples t-test showed that the experimental group has significantly higher scores ($M = 5.62$, $SD = 0.3011$) compared to control group ($M = 5.29$, $SD = 0.2998$), $t(9) = 7.3586$, $p < 0.001$.

Table 5.6: Paired-Samples T-test on the Data from the Second Peer Review Session

		<i>Control Group</i>	<i>Experimental Group</i>
Mean		5.29	5.62
SD		0.299815	0.301109
Mean difference	0.33		
p-value for paired t-test	4.29E-05		
t-value	7.358612		
Df	9		

2. What metrics can we use to measure the success of a mentoring relationship?

We did an explicit measure of the mentoring session by probing the participants on how much they thought they learned from the experience in both author and reviewer roles. They were also asked about their perception of the helpfulness of the peer review experiment. The results are presented in table 5.7.

Table 5.7: Helpfulness and Learning from the Peer Review Experiment

	<i>Control Group</i>		<i>Experimental Group</i>	
	Helpfulness	Learning	Helpfulness	Learning
Average	6.05	6	6.09	6.15
Standard deviation	1.11	1.41	0.93	1.10

We observed that participants from both the control and experimental groups perceived that they learned from the peer review system and also found the system to be helpful. However, the participants in the experimental group (6.09) found the peer review system to be more helpful than those in the control group (6.05). Also, their average learning (6.15) from the system is higher than the average learning of the control group (6). Their comments also showed that they liked the experience from the peer review session. One of the comments is stated below.

- *"It is very valuable because you can reach for excellence as it allows you to be corrected by your peers"* (study participant)

We also asked participants if they would be happy to recommend the peer review system to their schools. 72 (~98%) out of 74 participants agreed that it is a good idea that they would like to recommend to their schools. The other 2 participants that disagreed were from the control group and the comments that they gave are quoted below.

- *"It is very relative. In higher education, it is more difficult to install collaborative practices, since most of the teachers are hired and fees are not paid for that part of their work. No real policy of innovation in teacher training. Personally I find it useful, although the questions in front of each review were too oriented to a single type of audience and context"* (Study participant)
- *"There were no comments regarding my work that I could tap into my school. Again, teachers I interact with were not high school teachers. From another perspective, as the vice Chancellor for academic of a private school, I realized the lack of preparation of teachers in the basic cycle. It helped me understand the reality of the country"* (Study participant)

Their comments showed that they were not totally in disagreement with introducing the system to their schools, but they only envisage problems with their country's education policy to accommodate the system and also the preparedness of the teachers to introduce such collaborative task to their students.

5.4 Case study results and discussion for combined results

In section 5.3, we described the experiment with Chilean teachers in 2013. We could only report the results from the controlled experiment for the Mathematics teachers' group, due to a mix-up in the collation of data for the other two groups – Music and Language. However, we have the combined results for all the three groups – Mathematics, Music and Language. In this section, we set out to find answers to the following specific questions, from these results.

1. *Can we judge the quality of reviews by their lengths?*
2. *Does the back-evaluation of reviews encourage reciprocity of ratings in the peer review feedback?*

Below are the results obtained presented as they relate to the questions that we set out.

1. Can we judge the quality of reviews by their length?

Since the back-evaluation rating is the author's perception of the helpfulness of their reviews, we wanted to know if the review length affected these ratings. Table 5.8 shows the lengths of the

reviews and the correlation between the review length (measured by the number of words in it) and the back-evaluation ratings provided by the authors of the essays that were reviewed.

Table 5.8. Average lengths of the reviews & standard deviation; Correlation between back-evaluation & length of review; R-Squared (R^2)

	<i>1st Session</i>		<i>2nd Session</i>		<i>Correlation (r)</i>		<i>R-Squared (R^2)</i>	
	<i>Average</i>	<i>St.dev</i>	<i>Average</i>	<i>St.dev</i>	<i>1st Session</i>	<i>2nd Session</i>	<i>1st Session</i>	<i>2nd Session</i>
Mathematics	89.891	34.466	85.169	30.432	-0.094	-0.08	0.009	0.006
Music	108.667	35.487	103.395	37.521	-0.095	0.043	0.009	0.002
Language	99.453	32.05	83.289	41.984	-0.0305	-0.206	0.0009	0.042

First, we observed that the average review length got shorter in the 2nd session. Yet the standard deviation (St. dev) is rather high, so it is hard to speak of a trend. The correlation values vary and are mostly negative (i.e. longer reviews are rated lower), but nearly always close to 0. So for most reviewers, there is no correlation between their back-evaluation ratings and the number of words in their reviews. The one instance of weak negative correlation (-0.20591) for the language group in the 2nd session, was interesting to explore. We calculated R^2 in order to determine the impact of how well the review length could predict the back-evaluation ratings. The R-squared analysis results showed that the review length is a weak predictor of the back-evaluation ratings, with the highest value being $R^2=4.24\%$. Therefore, our results show that lengthy reviews were not perceived as being more helpful to the authors.

2. Does the back-evaluation encourage reciprocation of ratings?

Authors were encouraged to provide a back-evaluation of the reviews of their essay, given by the reviewers. Participants used pseudonyms in both peer review sessions, in order to mask their identities and also assign ratings from both peer review sessions to their pseudonyms. With the blind-review, we expected that authors would not try to reciprocate the ratings that the reviewers gave to their essay in the back-evaluation. To confirm our expectations, for each of the three groups, we ran a correlation test on the ratings given by the reviewers and the back-evaluation ratings given by the authors to the reviews they received (Table 5.9). If the authors were trying to reciprocate reviewers' good ratings and positive comments or retaliate the reviewers' low ratings and critical comments, we would find a positive correlation.

Table 5.9. Correlation between back-evaluation and review ratings (merging reviewers)

	<i>Correlation</i>		<i>R-Squared (R²)</i>	
	1st Session	2nd Session	1st Session	2nd Session
Mathematics	0.1899	-0.0331	0.0361	0.0011
Music	0.1808	0.2417	0.0327	0.0584
Language	-0.0161	-0.0331	0.0003	0.0011

Our results showed mixed weak correlations (-0.03 and +0.18) and one case with a slightly higher correlation value of 0.2417 in the 2nd session of the Music group. We calculated R^2 in order to determine how well the review feedback predicts the back-evaluation ratings. The results showed that the review feedback is a very weak predictor of the back-evaluation ratings, in the Mathematics and Music groups, with the highest value being $R^2=5.842\%$ for the 2nd session of the Music group. It seems that authors were not trying to reciprocate the reviews in the back-evaluation, but were only providing helpful and truthful feedback to their peers.

5.5 Conclusion

In this chapter, we described a large-scale experiment with teachers in Chile which aimed to validate if our modified peer-review approach with its two main components – the diversity-based peer-matching in groups for review assignment and the back-evaluation stage for authors to evaluate the reviews they received, is beneficial. We carried out a controlled experiment and used the obtained data to both answer the main question, stated above and to explore other interesting aspects of the peer-review mentorship process, such as whether reciprocation takes place between reviewers and authors, and whether the length of the reviews correspond to the ratings given by the authors of their helpfulness. We presented the results from the Mathematics teachers' group for the controlled experiment. We found that our participants were generally happy with the use of the peer review system in group peer mentorship. They found this system to be helpful in learning and about 98% of the participants indicated that they would be happy to introduce the system of peer review to their classes. However, the participants in the experimental group have a higher average for helpfulness and learning from the use of the modified peer review system than the control group. Participants in both the control and experimental groups showed some improvement in their writing (as shown in the average of the back-evaluation ratings of their review in table 5.5). However, the participants in the

experimental group had higher average peer ratings than the control group. The results of the paired-sample t-test conducted on these data showed that the difference is not significant. However, for the feedback on their peer reviews (back-evaluation ratings), we saw an improvement in the feedback of the participants in the experimental group over the participants in the control group, and the results of the paired-sample t-test conducted on these results showed that the difference is statistically significant. Although these results are not all significantly conclusive, we observed that the modified peer review process improved learning over the standard process, without diversity-based peer group assignment.

We could not get the results for the controlled experiment for the Music and Language groups, due to a data compilation error in which the original data for these two groups was lost by our Chilean colleagues and only aggregate data for the three groups was kept. Therefore, we were only able to present and discuss the combined results. We found in the literature that many approaches have been proposed to improve the peer review process. One of these approaches is the back-evaluation of peer reviews (Lan *et al.*, 2011; Adewoyin and Vassileva, 2014). However, little has been done on measuring the effectiveness of the back-evaluation in fulfilling the formative objective of peer review. Therefore, this study fills this gap by measuring satisfaction with the modified peer review process using the feedback from both authors and reviewers, checking if the back-evaluation phase encourages reciprocity/retaliation by authors to reviewers, and investigating the role of the length of the review as a predictor of the review quality and helpfulness. Our results suggest that providing opportunity for authors to back-evaluate the reviews of their work does not necessarily encourage reciprocation of ratings. Instead, we found that participants were just being helpful and honest with the feedback they provided as back-evaluation of their reviews. Also, our results indicate that the quality of the reviews, as perceived by the recipients, cannot be evaluated by their length.

CHAPTER 6

MATCHING PEERS IN GROUP PEER MENTORSHIP SYSTEM

In all the case studies discussed in chapters 4 and 5, peers were classified into peer review groups based on their capabilities, such that the weak peers were grouped with the strong peers following the criteria outlined in chapter three (see section 3.3.3). The groupings were done manually in these case studies, because the sample size was still manageable by a manual approach and the full implementation of the algorithm was not ready by the time of the experiment. However, following a manual approach would not be efficient in the case of a large number of peers. Therefore, we see the need to introduce a more advanced peer grouping algorithm that can handle large groups of peers with different levels of strengths, with respect to a particular dimension of competence (for example, writing, argumentation, coding). This chapter discusses this peer grouping algorithm, the results of the evaluation of this algorithm and its comparison with three other baseline algorithms used in group learning.

6.1 Problem Definition

The problem of assigning reviewers appropriately is essential in academic peer-review systems. Different systems provide different functionalities to support the review assignment, such as selection of competence areas by reviewers from a predefined list of areas, and selection of areas addressed by the submissions by authors. Thus the problem of review assignment is treated as semantic matching between the expertise of reviewers and the areas of the papers. Yet this approach has the shortcomings of any taxonomy-based matching approach – the predefined list of areas may not capture well the areas represented in the competences of the reviewers or the submissions, and both reviewers and authors may interpret the areas as they wish, which may lead to inconsistencies and sub-optimal matching. Also the lack of possibility to indicate strength of expertise in an area or degree of relevance leads to treating marginally relevant areas as equally important in computing the matching scores. To avoid these problems, an alternative approach, now implemented by most conference management systems is to allow reviewers to view the titles and abstracts of the submitted papers and bid on the papers they wish to review. This however, is often a very onerous task, and reviewers end up picking a few papers from the top of the list, while the papers down the list would not be selected by anyone.

The problem of reviewer assignment in a peer-review system for group peer mentorship is not so strongly dependent on the precise area of the submitted document and specific expertise of the reviewer. Usually, in group peer mentoring, the group of peers is focussed on one topic or question, and engage in discussion, argumentation, and providing criticism and constructive feedback, related to the individual contributions made by the participants. The main purpose is not quality evaluation and selection of individual contributions, but generally improving both the argumentation and discussion skills of peers in the role of mentees and the skills to provide constructive criticism and feedback for the peers in the role of mentors. The problem of reviewer assignment is transformed into the problem of how to group peers that provide feedback to each other to ensure that all peers benefit from the peer-review process, and everybody has something to offer. This problem is akin to the group formation problem in the area of CSCL. In section 6.1.1, we discuss some existing group formation algorithms in collaborative learning.

6.1.1 Group formation algorithms in collaborative learning

One of the basic grouping approaches employed by instructors is random grouping algorithm, which assigns students to groups with no particular pattern. However, a random grouping algorithm can result in unbalanced and ineffective group composition (Henry, 2013; Srba and Bielikova, 2015). Another approach is the stratified algorithm, which is an improved version of random algorithm that sorts students in the decreasing order of their capabilities (competence) and grouping is done starting from the top students to the weakest on the list. One drawback of this approach is that it will, in most cases; result in creating homogeneous groups of students, which are less efficient than the heterogeneous groups in yielding the desired learning outcomes (Manske *et al.*, 2015; Strnad and Guid, 2009; Wessner and Pfister, 2001). Researchers had proposed different group formation algorithms (see Table 2.2 for details) based on the traditional algorithms e.g. genetic, heuristic, random, clustering, optimization algorithms, multi-agent systems and group technology.

Wang *et al.* (2007) proposed a grouping algorithm called DIANA, which uses the psychological features of students to map them into heterogeneous groups using the traditional genetic algorithm. However, DIANA considered so many features that could make it too generic when determining groups of students. Also since it relies on students to self-report their features, its accuracy can be affected by error or bias in the students' response. In addition, authors were

silent on the scalability of DIANA. Moreno *et al.* (2012) also proposed a group formation algorithm based on genetic algorithm, to generate inter-homogeneous and intra-heterogeneous groups. However, the algorithm also relies on self-reported data by students in order to determine the characteristics that are used in grouping them. Yannibelli and Amandi (2012) also proposed crowding evolutionary algorithm that assigns students to nine different roles defined by Belbin (1983) and based on these roles, students are randomly placed in different groups. These groups reshuffled using mutation and crossover algorithms. The fitness level of each group formed is evaluated until there is an optimal combination of roles in each group, and the group formation algorithm stops. One drawback of this algorithm is that it relies on some random matching for the initial group formation. Also, the report was silent about how students' roles were determined for the initial random grouping.

Cavanaugh *et al.* (2004) also proposed TeamMaker, which is an instructor-focused grouping algorithm that uses both random and heuristic algorithms to map students into heterogeneous groups. Another similar algorithm is proposed by Christodoulopoulos and Papanikolaou (2007) as a web-based group formation tool, which relies on learners' features defined by the instructor to map learners into homogeneous and heterogeneous groups, depending on the preferences set by the instructor. However, they also rely on random grouping algorithm to generate heterogeneous groups. Tanimoto (2007) proposed the Squeaky wheel algorithm, which uses the students' ratings of their willingness to work with their peers to compute their mutual compatibility. Their mutual compatibility is therefore used to create heterogeneous groups of students. This algorithm relies on the initial ratings of students' willingness to work with their peers, which can result in homogeneous groups and some students, who might not be preferred by their peers end up being orphaned and might be grouped using some neutral default values.

Inaba *et al.* (2000) proposed a multi-agent system based algorithm, which models the learning goals of students as agents and negotiates between students' agents to form mutually beneficial groups. However, their report lacks the description of the architecture and evaluation of the system. So, we cannot confirm the scalability and effectiveness of the system over other existing algorithms. Soh *et al.* (2006) also proposed a multi-agent based group formation system called I-MINDS, which represents students and groups by intelligent agents that profile students

and the groups respectively. The students' agents negotiate with the groups' agents, based on their previous performances in the group activities. However, authors were silent about the source of data for the first group activities.

In 2015, Sarkar *et al.* proposed an automated group decomposition program, which implements k-means algorithm to classify students into heterogeneous groups. This program creates groups in stages. First, it groups students into initial homogeneous groups using their attributes (e.g. communication skill, fluency in the use of computer and in the group work). Then, students are re-grouped into heterogeneous groups using their knowledge of the subject. One drawback of this program is that the attributes used in forming the initial homogeneous groups are fixed and cannot be changed by instructors. Also, the system has not been evaluated in actual academic projects, and the source of data containing students' attributes is not defined. In 2012, Cocea and Magoulas (2012) proposed user behavior driven group formation tool. This tool uses case-based reasoning to model learners' behaviors, which also contain their learning strategies. The tool extracts their learning strategies to form the strategies-learners' matrix, which is passed to the clustering algorithm to create homogeneous groups of learners. Although, this tool implements a novel idea that considers learners' strategies, it does not address the specific benefits that learners derive from the groups that they are assigned to. Also, the tool generates homogeneous groups which are less efficient than heterogeneous groups in yielding the desired learning outcomes (Manske *et al.*, 2015). Liu *et al.* (2016) also proposed an algorithm based on the heuristic algorithm and uniform k-means clustering. This algorithm uses a cognitive diagnoses model named SDINA to automatically quantify students' skill proficiencies in binary value '0' or '1'. These quantified skill proficiencies are used to generate collaborative learning teams with dissimilar features. The system has only been tested with simulated and pre-defined students' data. In 2014, Aggrawal *et al.* proposed group formation algorithms for educational settings, which are based on modified clustering algorithm. These algorithms use learners' abilities in the subject of interest to determine their strengths. In this study, they assumed that each student is associated with certain ability. The algorithms then classify learners as "leader" or "follower". The word "leader" is used to describe a learner that has strong capability in the subject of interest, while "follower" describes a learner that has weak capability in the subject of interest. The algorithms map learners into groups in such a way that there are just few leaders in each group and each group has more followers than leaders, to ensure that leaders are spread

across groups and are able to maximize their skills to help followers. Unfortunately, these algorithms consider the learning needs of followers to be more important than the leaders. Also, since the algorithms have only been tested with simulated data with assumed abilities, further testing in classroom settings is still required.

6.1.2 Limitations of the existing group formation tools

The general limitations of the existing group formation tools / algorithms are discussed below.

1. **Limited constraints:** The existing systems model certain criteria, which are used in grouping learners. For example, DIANA uses a maximum of seven criteria (Wang, 2007). With this in place, the systems become inflexible and it becomes difficult for the users (instructors / learners) to include criteria outside the space of the criteria modeled in the system. Also, users have to strictly follow the listed criteria in the use of these existing systems, even if the criteria do not fully align with their requirements. In addition, some of the systems require some background information about the learners to do the grouping (Liu *et al.*, 2016; Vivacqua and Lieberman, 2000). This makes it difficult to use effectively when there is no sufficient information about learners, for example, in the case of new learners.
2. **Manual grouping to support the existing systems:** In certain instances, the criteria in grouping might result in “*orphaned learners*”, who are stranded learners that remain ungrouped (Tanimoto, 2007; Tobar *et al.*, 2004; Redmond, 2001). Therefore, the instructors will have to manually add them to certain groups, which might reduce the effectiveness of the group formation tool.
3. **Evaluation strategies:** Most of the existing grouping strategies rely on self-report by the learners, using questionnaire, to measure their effectiveness. However, research has shown that self-reports can be misleading and loaded with bias (Sorensen, 2008). Therefore, evaluation of a group-matching algorithm should be multi-faceted, with log and qualitative data from the system and observation reports from the instructors or teaching assistants used as backup to the self-reports from the learners.
4. **Goal of the grouping strategies:** Most grouping strategies discussed in literature, except few (e.g. Aggrawal *et al.*, 2014; Srba and Bielikova, 2015), are focussed on enhancing collaborative learning activities with no mention of the aggregate gains of peers, in terms of

learning or skill enhancement, from the collaboration; whereas the purpose of collaborative learning is to facilitate learning in order to enhance learners' skills.

6.2 Group formation as akin to assignment problem in Mathematics

We see the group matching problem in group peer mentorship as a bipartite graph. A bipartite graph is a "*special case of k -graph (where $k=2$) that the vertices can be decomposed into two disjoint sets such that no two vertices within the same set are adjacent*" (Weisstein, 2014). In the group peer mentorship matching, peers constitute both disjoint sets and can be matched to form mentorship group as in a bipartite graph, such that peer matching does not contradict the criteria set for matching. These criteria were discussed in section 3.3.3. We see potential in the algorithms used in solving assignment problems in mathematics, for example, the Hungarian algorithm, the stable marriage algorithm and sports tournaments algorithms (e.g. round robin, elimination, challenge, and knock out draw, 12 draw and king of the court).

6.2.1 The Hungarian Algorithm

The Hungarian algorithm is a combinatorial optimization algorithm that solves the assignment problem in polynomial time. It can be used to find the minimum assignment in situations like assigning workers to jobs or machines to tasks while still maintaining maximum effectiveness. To solve an assignment problem using Hungarian, the problem will be rephrased in terms of graph theory. The elements to be matched are seen as vertices in a bipartite graph. An adjacency matrix (with $n \times m$ elements) is derived from the bipartite graph where n and m are the sets of vertices from the bipartite graph. If the resulting matrix is not a square matrix, the matrix is augmented by adding a dummy row or column (as necessary) to make it a square matrix (Peltola, 2014). However, since we will be assigning people to people using a variant of the Hungarian algorithm, we will always have a square matrix. We have presented below the steps involved in using Hungarian algorithm to solve a sample assignment problem.

Example⁵: To assign four jobs (J1, J2, J3 and J4) to four men (W1, W2, W3 and W4) in such a way as to minimize the total cost of the assignment while maximizing workers' effectiveness. The cost of assignment of each man to a job is given as elements of the adjacency matrix below.

⁵ Example taken from <http://www.hungarianalgorithm.com/examplehungarianalgorithm.php>

	J1	J2	J3	J4
W1	82	83	69	92
W2	77	37	49	92
W3	11	69	5	86
W4	8	9	98	23

1. Start by subtracting the row minimum of each row from every element of the row. For example, we have to subtract 69 from row 1, 37 from row 2, 5 from row 3, and 8 from row 4. The resulting matrix is as follows:

	J1	J2	J3	J4
W1	13	14	0	23 (-69)
W2	40	0	12	55 (-37)
W3	6	64	0	81 (-5)
W4	0	1	90	15 (-8)

2. Next, we also subtract the column minimum from every element of that column. That is, subtract 0 from column 1, 0 from column 2, 0 from column 3 and 15 from column 4. The resulting matrix is given below.

	J1	J2	J3	J4
W1	13	14	0	8
W2	40	0	12	40
W3	6	64	0	66
W4	0	1	90	0

(-15)

3. We want to cover all zeros in the matrix with the minimum number of lines (both horizontal and vertical). In our case, there is a minimum of three lines, as shown below.

	J1	J2	J3	J4
W1	13	14	0	8
W2	40	0	12	40
W3	6	64	0	66
W4	0	1	90	0

x

4. Create additional zeros by subtracting the minimum uncovered element from all the other uncovered elements in the matrix. In this case, we want to subtract 6 from other uncovered elements.

	J1	J2	J3	J4
W1	7	8	0	2
W2	40	0	18	40
W3	0	58	0	60
W4	0	1	96	0

5. To test for optimality, we will cover all the zeros with the minimum number of lines, as we did in step 3. In this case, we have four lines. Since the number of lines equals the size of our matrix (n=4), an optimal assignment exists in the covered zeros and the algorithm stops. If there is no optimal assignment yet, delete the minimum element not covered by any line from each uncovered row, then add it to each covered column and repeat steps 3 and 4.

	J1	J2	J3	J4	
W1	7	8	0	2	x
W2	40	0	18	40	x
W3	0	58	0	60	x
W4	0	1	96	0	x

- a. The zeros highlighted in the following matrix show an optimal assignment in the matrix.

	J1	J2	J3	J4
W1	7	8	0	2
W2	40	0	18	40
W3	0	58	0	60
W4	0	1	96	0

This combination is highlighted in the original matrix below. It shows that W1 should perform J3, W2 should perform J2, W3 should perform J1 and W4 should perform J4 in order to have an optimal assignment with the cost value of 140 (That is, 69 + 37 + 11 + 23).

	J1	J2	J3	J4
W1	82	83	69	92
W2	77	37	49	92
W3	11	69	5	86
W4	8	9	98	23

Algorithm 6.1: Hungarian Algorithm⁶

⁶ Pertola, 2014; Bruff, 2005; <http://www.hungarianalgorithm.com/examplehungarianalgorithm.php>

An advantage of Hungarian algorithm is that it is polynomial. That is, it makes use of simple data structures and simple matrix operations (addition, subtraction, multiplication and non-negative integral operations) to obtain optimal solution (Frank, 2004). In short, the worst case computational complexity of Hungarian algorithm is $O(n^3)$.

6.2.2 Matching in the Stable Marriage Problem

We can also relate the group matching in group peer mentorship to the stable marriage problem, where peers represent men and women waiting to be matched to achieve stable marriages. A stable marriage problem, also an assignment problem, comprises instances of men and women waiting to be matched with their most preferred mate. Each person (male or female) has a list of preferences ordered (ranked) from 1 to n . Then, they are matched to their mates based on their preferences. A matching is only stable if there is no two people that would rather prefer each other to the mates they have been matched with (Iwama and Miyazaki, 2008). One of the algorithms used in solving the stable marriage problem is the Gale-Shapley algorithm (Gale and Shapley, 1962).

Gale-Shapley algorithm, also known as propose and reject algorithm (see the algorithm 6.2 below), was proposed in 1962 by Gale and Shapley, and it is guaranteed to achieve a stable matching.

```

" Initialize each person to be free
while (some man is free and hasn't proposed to every woman)
{
    Choose such a man  $m$ 
     $w =$  1st woman on  $m$ 's list to whom  $m$  has not yet proposed
        if ( $w$  is free)
            assign  $m$  and  $w$  to be matched
        else if ( $w$  prefers  $m$  to her mate  $m'$ )
            assign  $m$  and  $w$  to be matched , and  $m'$  to be free
        else
             $w$  rejects  $m$ 
}
"

```

Algorithm 6.2: Gale-Shapley Algorithm⁷

⁷ Gale and Shapley, 1962

In this algorithm, men propose to the women in the decreasing order of their preferences, so that once a man is matched to a woman, the woman is no more unmatched but she can trade up her mate. Whenever a man proposes to a woman, there are n^2 possible proposals, after which the algorithm terminates. The Gale-Shapley algorithm is nondeterministic. That is, the order at which the men propose to the women in the algorithm does not follow any pattern. However, the algorithm results in stable matching eventually. So, it does not matter in which order the proposals are done (Gusfield and Irving, 2003). However, a major drawback of this algorithm is that it favours one gender (male) over the other gender (female). Although the Gale-Shapley algorithm is computationally efficient and easy to implement, it is also easy to manipulate the algorithm since both genders are required to provide their preferred mates, which are considered in the matching (Pini *et al.*, 2009).

6.2.3 Tournament Organization

We can also liken the group matching problem in the group peer mentorship to the problem of sports tournament scheduling. A sports tournament comprises different teams that compete within a period of time, during which a winner will emerge. Some of the algorithms used in tournaments organization are the round robin, elimination and challenge. These algorithms require that the number of teams be known ahead of time, and also depend on the game type, time, facilities and equipment available (Ross, 2014; Krzyzanowski, 2014). In the round robin algorithm, each team is scheduled to play every other team at least once in the league. It is also used in the CPU process scheduling as a "*pre-emptive first come first served process scheduling*" (Krzyzanowski, 2014). Here, processes are queued up and allotted time slots to run in the CPU, when it is their turn. Processes run within their time allotments, after which they have to proceed to the end of the queue if they are unable to complete their run within their time allotments (Krzyzanowski, 2014; Ozdogan, 2011). In sports tournament, round robin involves assigning to each team a unique identifier and the following formulas are used to compute the number of games that need to be played per league, per team (entry) and per round (Ross, 2014).

1. Number of games per league = $N(N-1) \div 2$
2. Number of games per entry = $N-1$
3. Number of rounds:

- a. Even number of entries = $N-1$
 - b. Odd number of entries = N
4. Number of games per round:
- a. Even number of entries = $N/2$
 - b. Odd number of entries = $(N-1)/2$

Round robin scheduling requires that we know the number (N) of participating teams/entries beforehand. Also, it might result in an inefficient use of the available resources in terms of time and other facilities, since these are equally shared among the participating teams irrespective of the demand of each team (Ross, 2014; Krzyzanowski, 2014).

Another algorithm used in tournament scheduling is the elimination tournament. Elimination tournament refers to league tournament where teams have specific number of games to lose before they are kicked out of the tournament. There are single elimination (must lose one game to exit), double elimination (must lose two games to exit) and the triple elimination (must lose three games to exit) tournaments. Although the elimination tournament saves time by eliminating teams as quickly as possible in order to get a champion, it does not always produce the best team as the champion because the organization of the tournament could be noisy and teams might not be given fair chances before they are knocked out of the tournament (Ryvkin, 2005).

6.3 Proposed Matching Algorithm

Following the four constraints discussed in section 3.3.3, we proposed a matching algorithm, based on the Hungarian algorithm for assignment problem. We proposed to modify the Hungarian algorithm for the peer group matching for two reasons. As much as we would like to keep the constraints in view, every peer should be treated equally. This principle renders the stable marriage algorithm inapplicable because priority is given to a certain class of nodes over the other in the algorithm. The tournament organization algorithms require that the teams involved are known ahead of time in order to schedule their resource use in terms of time and other facilities. However, in this case, we require automatic matching of peers into groups. Therefore, we can only know the number of peers ahead of time and cannot predict the number

of groups that will result from our matching. Also, because we require peers to be treated equally, the Hungarian algorithm is more suitable.

To apply the Hungarian algorithm to an assignment/matching problem, the bipartite graph must be regular. In our case, we have a regular graph only if the number of weak peers (wp), identified from the initial calibration test, is not equal to or greater than half of the total number of peers. We present the pseudo-code for the modified algorithm below. This algorithm has been implemented using MATLAB.

1. Conduct a calibration test for peers
2. The calibration test will be graded by a senior peer. Based on their grades from the calibration test, assign peers into one of the three categories: weak peers wp , average peers ap and strong peers sp . [**constraint 2**: classify peers as wp , ap and sp].
 - A. wp has grade less than the average group grade
 - B. ap has grade equal to the average group grade
 - C. sp has grade greater than the average group grade
3. Cluster peers into smaller groups of four peers in each group. If they cannot be evenly grouped with four peers, the remaining peers should be treated the same way as the rest. [**constraint 1**: each group should not exceed four peers in number]
4. For each group:
 - a. If $\text{number}(wp) < (\text{number}(wp+ap+sp))/2$, go to c
[**constraint 3**: to ensure that peers form bipartite graph]
 - b. else if $\text{number}(wp) \geq (\text{number}(wp+ap+sp))/2$, then check
 - I. if $\text{number}(wp) = (\text{number}(wp+ap+sp))/2$, then (we will finish with very few matchings) do steps e to i
 - II. else if $\text{number}(wp) > (\text{number}(wp+ap+sp))/2$ go to step m
 - c. Check if $\text{number}(wp) \geq 2$, do steps e to i
 - d. else if $\text{number}(wp) < 2$, do e-f, then go to step j
 - e. Duplicate peers in each group and make them nodes in the two vertices of a bipartite graph
 - f. Match the each node from one vertex to another node in another vertex. No node (peer) should be matched to itself.

- g. Check if there is any wp to wp matching
 [constraint 4: a weak peer cannot review / mentor another weak peer]
 - I. If true, delete all wp to wp matchings
 - II. Else, proceed to step h
- h. Check if any wp to sp or wp to ap matching occurs more than once. That is, if any strong or average peer has more than one review from the weak peer
 [constraint 5: an ap or sp cannot receive review (be mentored) from more than one wp]
 - I. If true, discard all but the first matching with the combination wp to sp or wp to ap .
 - II. else proceed to step i
- i. Check if any matching has more than 3 peers on the list of the mentees
 [constraint 1: to prevent peers from being overwhelmed with work]
 - I. If true, trim down to 3 in each list and proceed to step h
 - II. else if false, proceed to step j
- j. Compute the peers' gain within each group
- k. Compute the time and space consumption
- l. Print out the final matchings, peers' gain, time and space consumption
- m. End program

Algorithm 6.3: Pseudocode for the proposed group matching algorithm

6.4 Sample peer matching from our algorithm

Given 5 peers and we need to match them in group peer mentorship using our group matching algorithm. If the number of weak peers (wp) is 1, we have a regular graph and we can proceed. In order to try a more difficult example, we present the sample matching with 5 peers, including 2 weak peers in the steps below.

1. Duplicate the peers into two vertices of a bipartite graph

wp_1	wp_1
wp_2	wp_2
ap	ap
sp_1	sp_1
sp_2	sp_2

2. Match each peer from each set to the peer on the second set. Eliminate the match from a peer to itself. The possible matchings are

wp_1 to (wp_2, sp_1, sp_2, ap)
 wp_2 to (wp_1, sp_1, sp_2, ap)
 sp_1 to (wp_1, wp_2, sp_2, ap)
 sp_2 to (wp_1, wp_2, sp_1, ap)
 ap to (wp_1, wp_2, sp_1, sp_2)

3. Considering our constraint 4, which says that a weak peer cannot review another weak peer, wp_1 cannot be matched to wp_2 and vice versa. Hence, the first two matchings are void - wp_1 to (wp_2, sp_1, sp_2, ap) and wp_2 to (wp_1, sp_1, sp_2, ap) . Since removing these two matchings will leave wp_1 and wp_2 with nothing to do, we will eliminate the weak peers (wp_1 or wp_2) from the first two matchings and the resulting matchings will be

wp_1 to (sp_1, sp_2, ap)
 wp_2 to (sp_1, sp_2, ap)
 sp_1 to (wp_1, wp_2, sp_2, ap)
 sp_2 to (wp_1, wp_2, sp_1, ap)
 ap to (wp_1, wp_2, sp_1, sp_2)

3. Considering our constraint 5, which says that a strong or average peer cannot receive review from more than one weak peer, we have to remove a matching to wp (if there is more than one wp on the list) in the first two matchings. The resulting matchings are:

wp_1 to (sp_1)
 wp_2 to (sp_2, ap)
 sp_1 to (wp_1, wp_2, sp_2, ap)
 sp_2 to (wp_1, wp_2, sp_1, ap)
 ap to (wp_1, wp_2, sp_1, sp_2)

4. Since we do not want peers to review or mentor more than 3 peers at once, we have to trim down the number of peers in the matchings in (3). So, we will end up with the following combinations.

wp_1 to (sp_1)

wp_2 to (sp_2, ap)

sp_1 to (wp_1, wp_2, sp_2)

sp_2 to (wp_1, wp_2, ap)

ap to (wp_1, wp_2, sp_1)

6.5 Manual Time and Space Analyses of our Algorithm

Let n be the number of peers. Steps 1, 2, and 3 of the algorithm represent the pre-processing. Each of these steps can be performed in time proportional to the number of peers (i.e. in linear time), $O(n)$. Likewise, steps 4a and 4b can be performed in time proportional to the size of the group, thus overall each step will take $O(n)$ time. Also steps 4c and 4d together will take linear overall time $O(n)$ although the timing of the Hungarian algorithm is cubic, $O(n^3)$ in the size of the graph. However, our graph here is of a small size. So, overall, even if we compute all the possible matchings for a certain group, or even for every group, the time is going to be linear $O(n)$. Checking constraint 3 in 4g takes time proportional to the size of the matching, which equals the size of the group, so again we have overall linear time, $O(n)$. Steps 4h and 4i are analogous to 4g and, therefore, take linear time each, $O(n)$. Printing the final assignment also takes linear time. In conclusion, the algorithm takes linear time to execute, $O(n)$, as long as the size of the group is constant.

The storage is also linear. Observe that even if we compute all possible matchings for a group of size s , we have storage of $O(s!)$, which is constant for constant s . Think of s as the size of the largest group. To store all these matchings for all $O(n/s) \in O(n)$ groups takes $O(s!n) \in O(n)$ time. Therefore, no additional storage is used by the algorithm.

6.6 Experimental Test of the Peer Matching Algorithm

In this section, we discuss the experimental evaluation of the peer-matching algorithm to compare its performances with the three other algorithms: random, stratified, and Aggrawal *et al.*'s (2014) algorithm. We chose to compare with these three algorithms for several reasons.

Random and stratified algorithms are the baseline grouping approaches used in most of the existing peer review systems and in classroom grouping for collaborative activities, while the third grouping approach is a recent algorithm and implements similar strategies with our algorithm with few exceptions. A random algorithm assigns students to groups randomly with no particular pattern followed, while a stratified algorithm first sorts students in the decreasing order of their abilities, then grouping is done with the top ability students considered first and put in the same group, followed by the next students on the list. In the benchmark algorithm by Aggrawal et al. (2014), peers are named leaders (*sp*) and followers (*wp*). One leader is allowed to stay in each group with the other group members as followers to ensure they help the followers. However, in our grouping approach, we want to ensure that each peer learns within their zone of proximal development. Thus, we limit the number of *wp* who can be matched into the same group with the *sp* (see section 3.3.3, criterion 3).

6.6.1 Datasets

Our datasets comprised a set of 1,080 synthetic peers with abilities represented as randomly sampled values from normal, uniform, and Pareto distributions, while we observed their performances when $n = 120, 360, 600, 840,$ and 1080 . Peers from the normal datasets were sampled with abilities from a normal distribution with mean 0 and standard deviation 1, while their abilities were sampled uniformly from (0,1) for the uniform datasets. For the Pareto datasets, peers' abilities were generated from the Pareto distribution with the shape parameter set to 3.

6.6.2 Evaluating the performance of the peer matching algorithm

We evaluated the performance of our algorithm based on three metrics - the value of the aggregate knowledge gain made by peers in each group formed, time spent to execute the algorithm, and memory space consumed while executing the algorithm. Since our peers were synthetic and there was no real interaction or problem solving in each group formed, we modeled knowledge gain using the linear equation

$$r_{i+1} = a * r_i + b_{i+1} \dots \dots \dots \text{Equation (6.1)}$$

where r_{i+1} refers to the new value of the knowledge level of each peer, r_i refers to the current knowledge level of peers. The difference between r_{i+1} and r_i gives the knowledge gain of peers

after each iteration of the algorithm. a is computed by finding the standard deviation of the current knowledge levels of all peers in the group ($a=std(r_i)$) and b_{i+1} is computed using the quadratic equation

$$b_{i+1} = a1(1) * b_i^2 + a1(2) \dots\dots\dots \text{Equation (6.2)}$$

Where $a1(1)$ and $a1(2)$ are two values selected from the three random numbers generated from the function $a1(3) = random(3,1)$. With this approach, we envisaged that there would be an increase in the knowledge gain by each peer, irrespective of the algorithm used, but we expected that there would be a difference in the absolute value of gain made by peers using the four grouping algorithms. Therefore, the values of the three performance metrics from our algorithm were compared with the values of the same performance metrics from three other algorithms – random, stratified and grouping algorithm by Aggarwal *et al.* (2014). For clarity, we will refer to the four algorithms as follows, and these terms will be used interchangeably in our subsequent writing.

Random algorithm: *algo1*

Stratified algorithm: *algo2*

Aggrawal et al.’s (2014) algorithm: *algo3*

Proposed algorithm: *algo4*

6.6.3 Aggregate Knowledge Gains by Peers

For every group formed using the four algorithms, we computed the aggregate knowledge gains of peers in each group formed from the algorithms, and these values are represented on Figure 6.1 (see data in Appendix A, part 1). For the uniform and normal datasets, we observed that all the four algorithms recorded a gain in the peers’ knowledge across the entire datasets, except for when the gains dropped in *algo1* and *algo2* for the normal datasets from 840. For the two datasets, our algorithm (*algo4*) had similar performance in peers’ knowledge gain as with *algo3*. However, for the Pareto datasets, our algorithm outperformed all the three algorithms, with the peak value when $n = 840$, from which point the total gain starts to diminish for both our algorithm and *algo3*. However, there was no gain observed for peers using *algo1* and *algo2*.

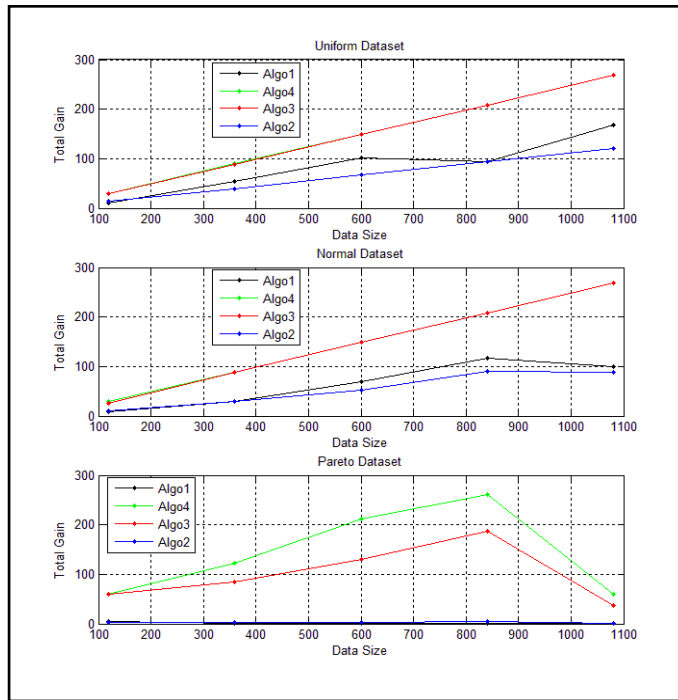


Figure 6.1: Total gain in peers` abilities from the three datasets (Four algorithms compared)

6.6.4 Time and Space Analysis

Here, we discuss the results of our test of the time and space complexities of our algorithm, which we compared with *algo1*, *algo2*, and *algo3* (Figs. 6.2 and 6.3; see data in Appendix A, part 2). We expected that our algorithm would keep the time and space usage to the minimum possible, when compared with the other three grouping algorithms.

As shown in Figure 6.2, we observed a drop in time for all the four algorithms with the uniform datasets up until when $n = 360$, and it remains stable until $n = 600$ when it starts to rise again. Also, in the normal and Pareto datasets, we observed a steady increase in time for all the algorithms with *algo3* having the highest rate of increase. However, *algo1* and our algorithm had the least time usage for all the datasets, and *algo3* had the highest time usage for all three datasets. On the overall, we observed that our algorithm had a linear time usage as we had shown in the manual analysis in section 6.5.

Figure 6.3 shows the space usage for the four algorithms. We observed that all the four algorithms have a rising space usage with an increase in the sample size for all three datasets. Our algorithm (*algo4*) still managed to outperform the other three algorithms with the least space

usage for the uniform datasets, the least space usage for the normal datasets up until $n = 600$, when it starts rising and eventually has the same space usage with *algo3*. However, it has the same space usage as the other three algorithms for the Pareto datasets.

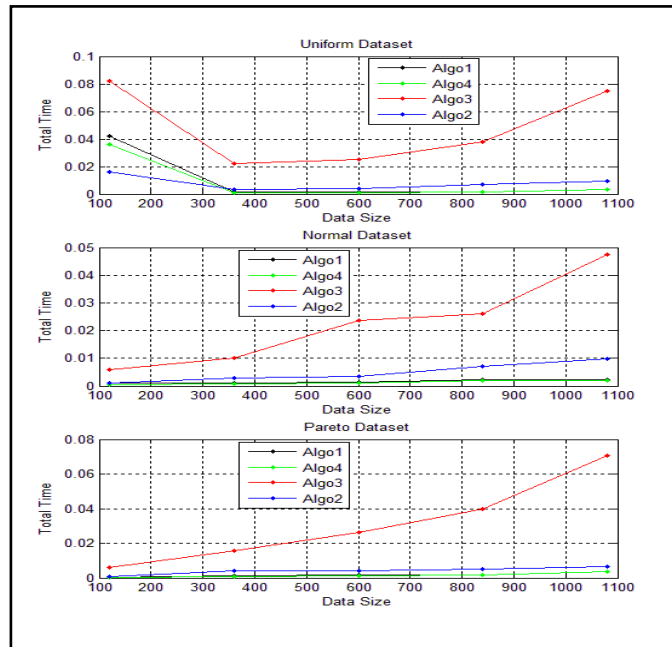


Figure 6.2: Time to run the four algorithms compared

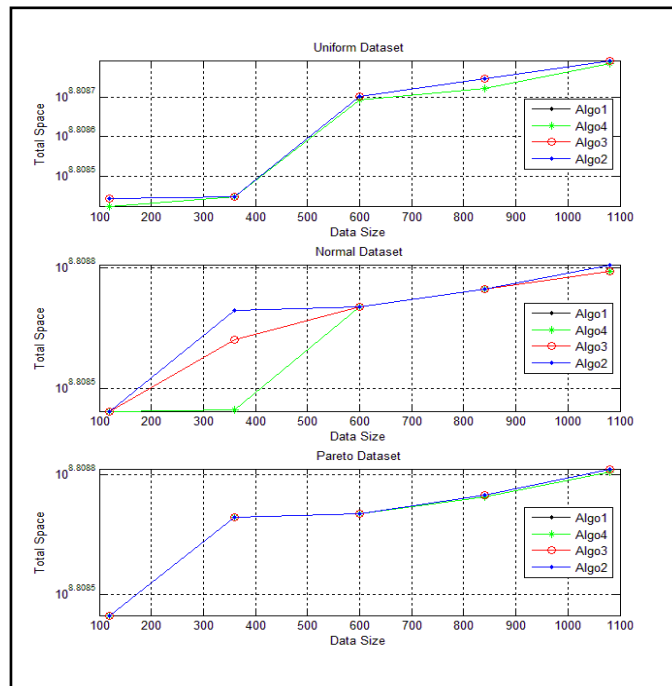


Figure 6.3: Space usage by the four algorithms compared

6.7 Limitations of our algorithm

Our algorithm has few limitations. First, we have restricted the number of peers in each group to four, in order to balance the workload on peers in each peer review session and also for easy coordination (Davis, 1993; Csernica *et al.*, 2002). However, in situations where we have more than four peers (e.g. 5) to be grouped, we will have to trim down the number of peers in order to keep to our criterion 1 (see the example in section 6.4). However, this can result in the trimmed peers having very limited feedback. Also in one of our grouping criteria, we proposed that peers be grouped based on their capabilities measured by an initial calibration task. In case where it is not possible to conduct an initial calibration task for peers, we have proposed the use of peers' past course average (see section 4.3.6). However, the past course average might not in most cases perfectly model the skill that peers are looking at improving using the modified peer review process. Also, the past course average cannot be used in studies involving professionals. In this case, the algorithm will rely on an initial random approach to grouping, and the peer ratings will be used for grouping in the subsequent peer review session.

6.8 Conclusion

Peer grouping is a crucial stage in group peer mentorship. A cohesive group is an indication of successful collaboration among the group members. Research had shown that a group with diverse skills will yield a positive learning outcome (Inaba *et al.*, 2001; Strnad and Guid, 2009; Wessner and Pfister, 2001). In this chapter, we have proposed a group-matching algorithm that relies on an initial calibration test depending on the purpose of the grouping to initiate peer grouping. In our experiments, we used an initial essay because the purpose of the group peer mentorship in our case is to improve the writing and argumentative skills of peers. The subsequent groupings are based on the outcome of the previous group interactions. Our approach is based on the Hungarian algorithm because it is easy to implement, does not require a prior knowledge of group components, does not give priority to any group member, and is not subjected to manipulation unlike the Gale-Shapley algorithm (stable marriage algorithm), which can be manipulated to give preference to a particular gender or peer. While the traditional Hungarian algorithm can be computed in a polynomial time $O(n^3)$, our algorithm can be computed in linear time $O(n)$, which offers better performance.

We ran experiments with 1,080 synthetic peers with abilities sampled from the uniform, normal, and Pareto distributions. With these datasets, we observed peers' knowledge gain separately for the groups formed by the four algorithms: random, stratified, Aggrawal et al.'s (2014) algorithm, and our group-matching algorithm. Considering the performance in the three datasets, our algorithm records the highest aggregate knowledge gains for peers on the overall for the three datasets. Also, for the time and space analyses, our algorithm has the relatively best performance. Overall, when we consider the three factors, peers' knowledge gain, space, and time, our algorithm has a better performance than *algo3*, which lags in time usage.

CHAPTER 7

ALTERNATIVE TOOL TO SUPPORT GROUP PEER MENTORSHIP – USE OF WIKI

While exploring the different collaborative learning tools that could support group peer mentorship, we implemented a modified wiki system. In this chapter, we present the description of the wiki system and the results of the experiment that we conducted with the undergraduate students of Computer Science while using the wiki system for their course work.

7.1 Introduction

Wikis are web pages that allow users to add, modify or delete contents, in collaboration with other users (Wikipedia, 2015). Wikipedia is the most famous application of wiki technology, ranking 7th on the Alexa's ranking of websites and with 1.2 billion unique visitors (Wikipedia, 2014). Use of web 2.0 technologies like wikis in education reflects a shift in the education paradigm from lecture and individual homework-based to a paradigm emphasizing student engagement and the construction of knowledge through collaboration and peer-help, which according to the socio-constructivists are powerful sources of learning (Piaget, 1928). Existing research on how wiki can enhance students' collaboration and learning are inconclusive in their findings (Judd *et al.*, 2010; Leung and Chu, 2009; Engstrom and Jewett, 2005). While some researchers found that wikis possess features that would be of great benefit to collaborative learning when an apt attention is given to sound pedagogy in its implementation (Engstrom and Jewett, 2005), others found that wikis do not necessarily encourage collaboration among students (Leung and Chu, 2009). Hence, the benefits of wiki in supporting collaborative learning still require further exploration with regards to whether it supports learning and collaboration. In this research, we used a modified wiki to support collaborative essay writing in a senior undergraduate Computer Science class, in order to discover whether it supports learning of writing and argumentation skills, and whether the students are motivated to use it and see it as a useful learning tool.

7.2 Research Tool

We developed a research tool, a wiki system called *WikiMentor*. *WikiMentor* is a customized MediaWiki system, which notifies users when the text they entered has been edited by somebody

else, and lets them review the change, accept or reject it and comment on it. Unlike regular wiki, it requires users to log in and provide an email address to which notifications can be sent. By comparing the content differences between every former and latter revision of an article, *WikiMentor* is able to figure out the authorship of each character (or word, sentence) and send email notification to the authors when their contents have been modified by others (see Figure 7.1). These email notifications only inform the authors (editors) that their contents have been changed.

A dialog box (see Figure 7.2) was added in each wiki article page. The dialog box is triggered when the user revisits a page after she has edited some texts and there have been subsequent edits, by others, of their texts. Two sub-functions are embedded in the dialog box, namely *content changes* and *acceptance/evaluation of changes*. '*Content changes*' lists all modifications of an article made between the newest version and the latest version contributed by the user in reverse chronological order. It further helps users to locate the modifications by highlighting the added (or deleted) content in the original place of the text. In this way, the users become aware of every change made to the article after the last time they logged into the system and made edits. For each change of content contributed by the user, there is the function ‘‘*Accept change*‘‘, which allows the user to accept or reject the change. The acceptance is considered equivalent to a positive rating (+1), and rejection – to negative rating (-1). In this way the system collects the user’s opinions of the quality of other users’ work on the text which really matters to the user (her own contributed text). We compute the user’s reputation from the acceptance or rejection, collected from the different users whose contributions were edited by the user using the formula below (Noorian *et al.*, 2011).

$$T = (r+1)/(r+s+2) \dots\dots\dots \text{Equation (7.1)}$$

Where T refers to reputation, r refers to the number of positive ratings (accepted changes), while s refers to the number of negative ratings (rejected changes).

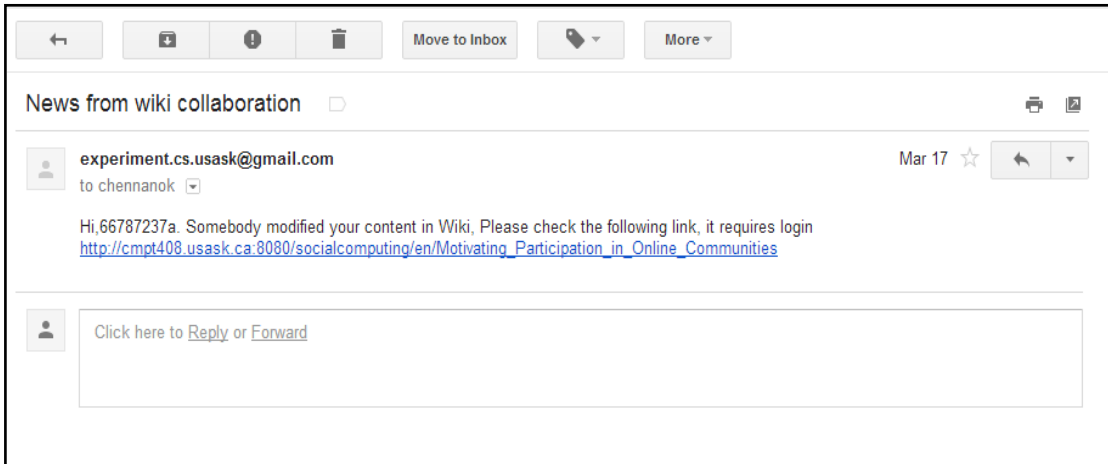


Figure 7.1: Sample email notification

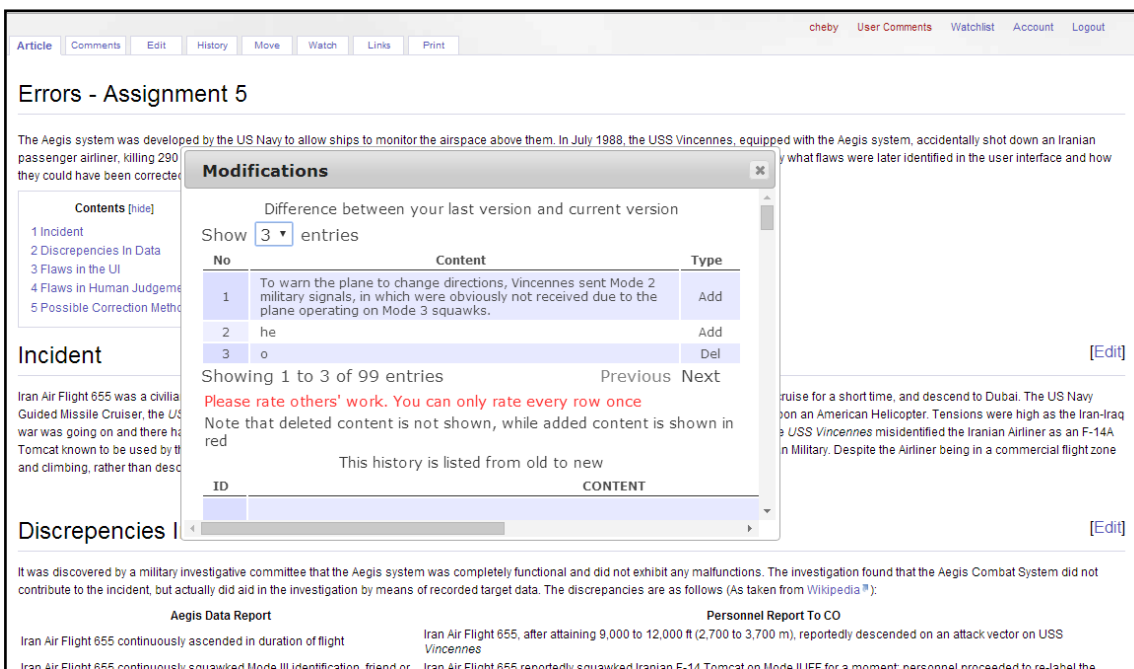


Figure 7.2: Dialogue box showing content changes & acceptance/evaluation of changes

In summary, in addition to the generic features of a wiki system, we included the following features to *WikiMentor*:

1. Email notification when changes are made to editor's contribution
2. Highlighting changes made to every editor's contribution

3. Opportunity to accept or deny changes made to their contribution
4. Computation of their reputation based on the acceptance or denial of their contribution or changes. While the reputation value was not visible to the editors, it was used in the analysis to check the editors' responses with the system data.

7.3 Experiment

This experiment was conducted in parallel with the peer review experiment in ethics 2014 class, with the same 10 participants from the peer review experiment. The participants started with two peer review sessions, then they engaged in a collaborative writing session using wiki, followed by another peer review session. Following this sequence, the participants engaged in a total of four collaborative writing sessions required for their coursework using *WikiMentor*. The wiki editors in our case were the students. The coursework required the students to write collaboratively essays on different topics assigned by the instructor each week. The students created the wiki page dedicated for the weekly assignment. There was no designated author responsible for each wiki page. The students were encouraged to contribute to the wiki assigned each week using pseudonyms. Students could add new contents, edit and delete the existing contents of the wiki. In order to ensure that students make distinct and meaningful contribution, their contributions to each wiki article were graded by a designated teaching assistant (TA), who is a senior graduate student and has taught the class as Sessional instructor. We only had one marker in order to prevent increase in the cost of the experiment since the TAs (markers) would be paid and also to avoid complexity that might arise from giving conflicting feedback to the students from different markers. The grading was done by assigning one grade for the final article and then deviations of this mark (both positive and negative within 15%) were assigned to individual students based on how substantial was their individual contribution, judged by the TA. In this way, we aimed to create positive interdependence among the students and as well enforce both group and individual responsibilities, since they all know that not only does the entire group contribution matter, but their individual contributions also count towards their final grades. In order to mitigate the subjectivity of the marks given by the TA, the grading of their final article was based on three criteria with some weights assigned to each criterion,

- 1) issues raised, weight 0.3;

- 2) completeness and logic of the argumentation, weight 0.4; and
- 3) writing style and grammar, weight 0.3.

Students had seven days to make contributions to each wiki article, after which the article was locked and grading started.

For every edit made to their contribution, each student got notified by email and the resulting changes were highlighted within the individual interface of the wiki system. Therefore, the user could either accept or reject the changes and this translated into a rating value of the change, that could be either positive (+1/accept) or negative (-1/deny), and was used in computing the reputation of the student who did the change as described in the previous section. We did not reveal the calculated reputation values to prevent the students from cheating or gaming the system. However, they were aware that their edits to others' contributions would either be accepted or rejected.

For each participant, we collected data on the individual weekly contribution quality (WCQ). Their individual WCQ is the number of their contributed characters that survived revision by the other participants, which is calculated using the formula:

$$WCQ_i = (\#characters\ owned\ by\ i\ in\ the\ final\ version) / (\#characters\ contributed\ by\ i\ over\ the\ week\ in\ total).$$

We also collected data on the revisions that they made, the time they spent making their contributions and revisions, and the numbers of their revisions that were accepted or rejected by the authors. Also, we kept history of their contributions and revisions, which could be viewed from the “history” tab once they logged in to the wiki system. At the end of the term, participants were given an exit questionnaire to evaluate their experience.

7.4 Results and Discussion

1. Does collaborative writing using wiki help students to improve their writing skill?

The participants engaged in collaborative writing using the wiki for 4 weeks. There was no designated group leader or author. Therefore, anybody could start each wiki article while others joined in adding more texts. At the end of each weekly article, we sent them the grades assigned to both their final article and their individual contributions, by the TA. These grades were used to ensure group and individual responsibilities (Barkley *et al.*, 2005). Grades assigned to the final

articles over the four weeks are shown in figure 7.3. We found that there was a positive improvement in their grades from 75% in the first week to 90% in the fourth week.

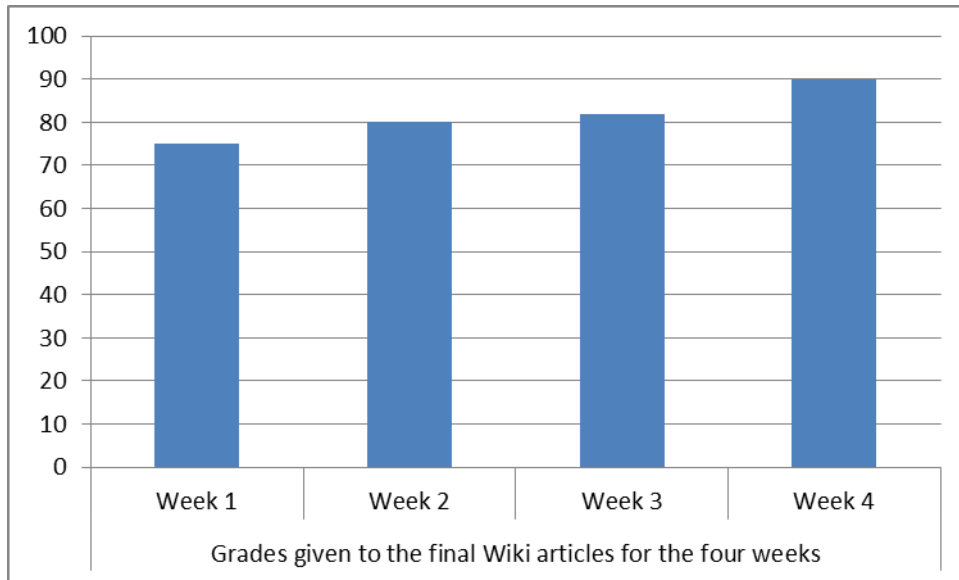


Figure 7.3: Grades given to the final articles for the four weeks

We also found a growing trend in their average individual grades over the four weeks from 78.60% in the first week to 91.89% in the fourth week (Table 7.1).

Table 7.1: Grades assigned to individual contributions

<i>Users</i>	<i>Week1</i>	<i>Week2</i>	<i>Week3</i>	<i>Week4</i>
A	75.00	90.00	87.00	95.00
B	80.00	80.00	82.00	95.00
C	77.00	80.00	82.00	85.00
D	85.00	85.00	94.00	93.00
E	75.00	75.00	77.00	87.00
F	77.00	80.00	90.00	90.00
G	75.00	90.00	100.00	100.00
H	85.00	90.00	77.00	97.00
I	77.00	-	75.00	-
J	80.00	75.00	77.00	85.00
Average Grade	78.60	82.78	84.10	91.89

We collected data on the individual weekly contribution quality (WCQ) for each student (results in Table 7.2). The results show a growing trend in the quality of the weekly contributions

of some of the participants, except in week 1 when they mostly had very high weekly contribution quality. One reason that can be attributed to these high values is that majority of the students gave their contributions towards the deadline, when it was practically impossible for the other students to edit their contributions. The result also shows a decline from a high class average in week 1 to a lower class average in week 2. This can be attributed to the last minute contributions made by the students, which had a huge impact on the overall class average. However, there was an increase in the WCQ for each student in the subsequent weeks, except in the week 4, when the values dipped lower, still a good trend on the overall.

Table 7.2: Weekly contribution quality for each student

<i>User</i>	<i>Week 1</i>	<i>Week 2</i>	<i>Week 3</i>	<i>Week 4</i>
A	0.9895	0.9058	0.9975	1
B	No contribution	0.5527	0.6967	0.9422
C	1	0.993	0.908	1
D	0.8201	0.9981	0.9336	0.9993
E	1	0.7008	0.9545	0.9282
F	1	1	0.9778	0.911
G	0.9523	0.7222	0.9973	0.9501
H	0.9994	0.9911	1	1
I	1	No contribution	0.9846	No contribution
J	0.9991	0.7897	0.9925	0.332
Average	0.9734	0.8504	0.9443	0.8959

2. What is the students' perception of the use of wikis for collaborative writing?

Participants were given an exit questionnaire. The exit questionnaire contained questions aimed at sampling their opinions about the competence of other contributors, the helpfulness of the wiki system in improving their writing, their satisfaction with the wiki system and the motivational strategies, which we included in the wiki system, that helped their writing and learning. We received 10 responses, though few of them abstained from answering certain questions.

We asked questions about their general impression of the other contributors to the wiki articles. The options given and the summary of their responses are presented in table 7.3.

Table 7.3: Participants' perception of other contributors

<i>Options</i>	<i>Number of respondents (out of 10)</i>
Competent	6
Provided detailed contribution	6
Helpful	6
Not helpful	1
Not competent	1
Lacked substance	1

Participants found other contributors to be competent, detailed and helpful on the average, which is also confirmed in their comments. e.g.

- *"I found sometimes people were too thorough, and by the time I went to record my thoughts, everything I wanted to say had already been said."* (Study participant)

We asked to know if they were actually comfortable with using wiki system and 80% said they were very comfortable. Although they used pseudonyms while contributing to the wiki articles and they expressed that they were comfortable with using wiki, we discovered from their comments that a few of them had reservations for editing other contributors' contents, while some did not have problem with this. Some of their comments are quoted below:

- *"It was uncomfortable deleting other people's work, but I was comfortable expressing my own ideas"* (Study participant)
- *"... I have no reservations about editing."* (Study participant)

Also, 60% of the participants objected to the further use of wiki in their coursework.

Some of the reasons that they gave are quoted below.

- *"...the wiki was a forced exercise in frustration, boredom, and annoyance"* (Study participant)
- *"With the wiki, what I wanted to say was often already said by someone else"* (Study participant)

Since both peer review and wiki systems were used in the class, we asked our participants about their preference between the two systems to support mentorship. 90% answered that they preferred the peer review system. This shows that, although there was a noticeable improvement in their writing, participants did not like to use Wiki in their collaborative writing, because they see it as an enforced and boring exercise.

We found out that participants found other contributors to be competent, helpful and detailed in their contributions. Although, 80% of participants feel they were comfortable with using the wiki system, we found out that some of them still hold some reservation for editing other contributors' contents. Many factors could have contributed to this that we hope to find out in our future work. Despite their positive attitude towards other contributors and their expressed competence in the use of wiki, participants were generally not happy with the further use of wiki in their collaborative writing.

3. What strategies can be used to improve students' participation in collaborative writing using wiki?

We asked participants about the features of the wiki system that motivated them to keep participating in the collaborative writing process. Participants could select as many features as were applicable to them on the list. See table 7.4 for the summary of the options and the selection made by the participants.

Table 7.4: Participants' preference for motivation strategies

<i>Options</i>	<i>Number of respondents (out of 10)</i>
The open contribution format of collaboration	6
Free writing style in wiki	6
The email notification when changes were made to their contents,	3
The highlighting of the changes made to their contents,	4
Their perceived status to their peers	7
Use of pseudonyms	2
Marks given by the TA	6

As shown in table 7.4, only four of the options were chosen by more than five participants. These are "The open contribution format of collaboration" (6), "Free writing style in wiki" (6), "The perceived status to their peers" (7) and "Marks given by the TA" (6). These results showed that wiki editors actually required motivation, even if used for educational purpose. Their perceived status was borne out of their reputation score, which was not visible to the participants, but computed from the acceptance or rejection of their contribution to the wiki articles. This showed that participants care about the acceptance and rejection of their contributions by other participants, and the awareness of such feature could motivate them to

ensure that they always make reasonable contribution. Also, participants got to see both their group and individual marks from every weekly article, before they moved on to the next wiki article. The results here showed that they were motivated by the feedback from the marks and the marker's comments that they received weekly from their contributions. Overall, students are generally motivated by approval or feedback from their peers and an authority figure, in this case, the teaching assistant (TA). Two selected options, free writing style and open contribution, are characteristic of any wiki system, which corroborates the findings by (Engstrom and Jewett, 2005) that the ease of use of wiki makes it a valuable tool in collaborative learning. The other two popular options (perceived status to their peers and marks given by the TA) correspond to enhanced performance feedback that can be usefully and easily incorporated in educational wiki systems. It was disappointing to find that the other new features introduced in our wiki, aiming to increase awareness of peer-feedback (the email notification when changes were made to their contents, the highlighting of the changes made to their contents, emphasizing fact that other editors are accepting or rejecting their contents, and the use of pseudonyms to encourage students to be more critical) were not popular options chosen as motivating students to participate in the system. Yet the students could only be aware of their reputation by being aware of the accepted changes that they made to their classmates' contributions. So it is possible that in answering this question, they focused on the higher motives for participation, assuming that the others (not selected items) are just technical means to achieve them. Further research will seek to clarify this issue.

7.5 Conclusion

A lot of research that had been conducted on wiki and collaborative learning were mostly targeted at whether wiki actually encourages collaboration or not. However, in this research, we studied the effects of the use of wiki as a source of peer feedback on improving writing skill and helping participants in collaborative learning. We also looked at the motivation strategies that can trigger meaningful contributions to wiki. The participants felt that wiki was not really helpful and it was rather a boring activity; however, the quality of their writing improved, both measured by the proportion of contributed text that remained in the final article and the grades assigned by the marker (TA) based on the quality of the collaboratively written article and the individual contributions to it. The findings from this study show that the wiki writing exercises are helpful

in improving students writing skill and that self-report from users is not enough when measuring cognitive and affective states. Our results also suggest that students require extrinsic motivation in form of feedback from their peers and an authority figure (e.g. instructor, TA) to enhance their quality of contribution to wiki.

Despite the limitations in the sample size of this study and the reliability of the ratings given by the marker, this study is pivotal for teachers, who will like to deploy wiki for collaborative writing in their classrooms. The findings here show that students require extrinsic motivation to participate in wiki writing. Therefore, teachers should consider the use of appropriate motivational features while deploying wikis, and that they should not only rely on the feedback from the students, but also feedback from an authoritative person (TA or marker), to measure the effectiveness of the use of wiki in collaborative writing and learning.

CHAPTER 8

CONTRIBUTIONS AND CONCLUSION

The goal of this research is to support peers for the purpose of group peer mentorship using a modified peer review framework. Group peer mentorship brings peers together for support and learning in groups to achieve their academic, career, or psychosocial goals (Ragins & Kram, 2007). In the literature review section, we identified two collaborative learning tools— Wiki and the peer review system—that could support group peer mentorship. Studies using both systems showed that peers preferred the peer review system to the Wiki system. However, we found that the features of the traditional peer review system are necessary but not sufficient to support the mentorship model. Therefore, we proposed a modified peer review framework, which was used in the six studies described in Chapters 4 and 5. We have discussed how the findings from these studies answer our research questions below.

8.1 Discussion of Our Findings

To support group peer mentorship using our modified peer review framework, we proposed four research questions (see chapter 3):

RQ1: How is the group composed for each peer review session?

RQ2: Are incentives required for providing high-quality feedback and, if yes, what incentives?

RQ3: What metrics can we use to measure the success of the mentoring relationship?

RQ4: What factors can influence peers' feedback in the mentoring relationship?

Please find in the subsequent sections our findings for the four research questions.

8.1.1 Research Question 1

To answer RQ1 on how to compose each peer review group, we proposed that peers work in small heterogeneous groups of at most four in each group. In each group, we proposed having a mix of weak and strong peers following our five peer-matching criteria described in section 3.3. That is, the number of peers should not exceed four in each group. The number of weak peers in

each group should be less than half the total number of other peers in the group to ensure that the weak peers learn from the strong peers and also to ensure that the strong peers are not overwhelmed with only weak feedback. We want to ensure that all peers learn within their zones of proximal development.

We conducted six studies over 4 years, four with undergraduate and graduate students of computer science, one with teachers in Chile, and one with masters of professional accounting students at the Edwards School of Business. In the four peer review studies with the graduate and undergraduate students of computer science, we explored the participants' perceptions of the peer review process by asking for their level of agreement with some positive and negative statements. The results showed a relatively high average for levels of agreement with the positive statements and a relatively low average for levels of agreement with the negative statements. In the controlled study with the teachers in Chile, participants were classified into control and experimental groups. Participants in the experimental group followed our peer-grouping approach, whereas participants in the control group were grouped randomly for the entire peer review study. We observed that the participants in the experimental group had significantly higher average scores from their back-evaluation than the participants in the control group. So those in the experimental group received better peer-reviews than those in the control group. However, we weren't able to show the development of better writing skills in the experimental group. Although the participants in the experimental group had a higher average ($M = 5.63$, $SD = 0.45$) in their essay feedback (review) than those in the control group ($M = 5.33$, $SD = 0.25$), a further analysis using paired samples showed that the results were not statistically significantly different [$t(2) = 2.60$, $p = 0.12$].

8.1.2 Research Question 2

To answer RQ2, which asked whether incentives were required to provide high-quality feedback, we proposed both intrinsic and extrinsic motivations for peers.

1. As an intrinsic motivation, we proposed that peers give back-evaluation of the helpfulness of the reviews they received and that peer mentors also give feedback of their learning from the peer review process. In sections 4.3.1 and 4.3.2, our results showed that peers were motivated to give thorough and quality feedback as a result of the back-evaluation of their feedback by their peers. In fact, they mentioned that they preferred

their peer ratings to the feedback given by the dedicated marker. In a similar study with undergraduate students in 2016, our results showed that the participants had a high regard for the feedback they received from their peers, so much so that 75% of them decided to voluntarily participate in peer review of their term papers. However, in the 2014 study with the research and methods class (CMPT 880/890, section 4.3.3), participants mentioned altruism (being helpful to their peers) as the factor that motivated them to give thorough and quality feedback, albeit this finding is still open to further research.

2. Also as a form of extrinsic motivation, we proposed that peers use pseudonyms for the peer review in order to not be afraid to give honest, critical, thorough and helpful feedback and also to be able to accrue their peer ratings, which could constitute their reputations in the community of peers. In the studies described in sections 4.3.1 and 4.3.3, we tested for anonymity and its effects on the quality of essays and reviews produced by the participants over time. We found that anonymity encouraged peers to be critical in their reviews, which translated to increasing the quality and helpfulness of their reviews. In the study presented in Chapter 5, the use of pseudonyms enabled us to accumulate participants' peer ratings and assign these ratings to their pseudonymous identities, which was useful in computing the relationship between the ratings they gave their peers and the back-evaluation ratings they got in return from their peers, in section 5.4.

8.1.3 Research Question 3

To answer RQ3, which asked for the metrics that can be used to measure the success of the mentoring relationship, we proposed that both peer mentors and mentees give their feedback with respect to their learning and the helpfulness of the peer mentorship session, because the main objective of introducing the group peer mentorship, which we have implemented as a modified peer review system, is to help mentees improve their writing skills and mentors – their critique skills. Although the learning outcomes measured by their grades did not show a steady pattern of positive change in sections 4.3.1 and 4.3.2, the participants' responses to the questionnaires confirmed that they perceived they had learned and improved their skills as peer mentors (reviewers) and mentees (authors). Also in section 4.3.5, we observed a steady increase

in the average essay scores of participants when they engaged in the peer review study and a drop in their grades in the week when they did not engage in peer review of their essays.

8.1.4 Research Question 4

To answer RQ4 about the factors that can influence peers' feedback in the mentoring relationship, we collated some data about peer ratings of mentors and mentees and their interpersonal trust scores. To find the effect(s) of peer ratings on their feedback, we calculated the correlation between review ratings given by the peer mentors and the corresponding back-evaluation ratings provided by the mentees. Our results showed that the review feedback given by peer mentors was a weak predictor of the back-evaluation ratings they received from their mentees. That is, mentees were not trying to reciprocate the reviews in the back-evaluation but were only providing helpful and truthful feedback to their peer mentors.

To find the effect of interpersonal trust on peer feedback, we calculated (i) the correlation between peers' interpersonal trust scores and the review ratings they gave their mentees as peer mentors, and (ii) the correlation between their interpersonal trust scores and the back-evaluation ratings they gave their peer mentors as mentees. We calculated their interpersonal trust scores using the Rotter trust/distrust scale. The results showed that the peers' rating behaviour was influenced by their trust score, and this depended on the role they assumed in the mentoring relationship. For example, we saw a moderate negative correlation between mentees' interpersonal trust and the back-evaluation ratings they gave their peer mentors, whereas we found a weak positive correlation between mentors' interpersonal trust scores and the review ratings they gave their mentees.

We also checked whether we could judge the quality of reviews by their length in words by calculating the mentors' review lengths and the corresponding back-evaluation ratings they received from their mentees. The back-evaluation ratings are the perception of helpfulness and quality of the reviews by the mentees. Our results showed that lengthy reviews were not necessarily perceived as being more helpful or having higher quality by the mentees.

8.2 Contributions

This dissertation contributes not only to mentorship, it contributes to the area of collaborative learning. The specific contributions can be classified into four groups, which are discussed below.

1. Group composition for collaborative learning.

We reviewed some existing algorithms, which are used for grouping peers for collaborative learning. However, these algorithms are limited by the specific constraints in their implementation, use of manual grouping, and evaluation strategies that are susceptible to noise. We proposed a group-matching algorithm based on the principles of the traditional Hungarian algorithm, following our grouping constraints. This algorithm was manually tested in five of our peer review studies. A simplified version was implemented for the Chilean study, because the algorithm was not fully developed at that time and there was very short time available to implement and embed the full version of the algorithm in their system before the teachers' seminar where the study was conducted. However, we evaluated the full version of the algorithm with 1,080 randomly sampled data from uniform, normal, and Pareto distributions using MATLAB. We compared the performances of our algorithm in terms of knowledge gain of peers from the group formed, and time and space consumption of the algorithm with the random, stratified, and group algorithms by Aggrawal *et al.* (2014). Our algorithm demonstrated good performance with more gain in peers' knowledge from the group formed, and it showed the best time and space consumption on the overall. With these results, we have identified four important factors for optimum group formation for collaborative learning:

1. Identification of skill levels of peers to ensure they receive appropriate help
2. Heterogeneous groups to enable transfer of skills within the community of peers in each group
3. Small groups to balance the workload on peers and enable them to give quality and helpful feedback to their peers
4. Periodic evaluation of total gains in every group formed to ensure that learning is taking place

2. Motivation strategies to encourage quality feedback in collaborative learning.

As an extrinsic motivation, we introduced the concept of back-evaluation, a process whereby authors evaluate the quality and helpfulness of the reviews they receive from their peer reviewers. Our findings showed that back-evaluation of reviews by the authors encouraged reviewers to give thorough and helpful reviews. There are two key objectives of peer review—summative (to assess the quality of scholarly work) and formative (to provide constructive feedback that can help peers to improve their skills). Although the traditional peer review system encourages reviewers to give feedback to help authors improve their work, it does not give room for reviewers to receive feedback from the recipients of the reviews they give (authors). With back-evaluation in place, reviewers are encouraged to provide more critical and helpful reviews and eventually become better reviewers. Therefore, the peer review process can provide formative feedback to help both reviewers and authors improve their skills.

Also as an intrinsic motivation for peers to give constructively critical and helpful feedback, we encouraged our participants to be anonymous, but still maintain an identity by using pseudonyms. With pseudonyms, they were able to preserve their peer ratings over time. Our findings showed that anonymity enhanced the quality of the collaborative venture by allowing peer reviewers to give thorough, critical, and unbiased reviews of the authors. As an extrinsic motivation, pseudonyms were used to preserve peers' ratings in their communities. In this research, these ratings were used to conduct some further investigation on reciprocation of ratings and dynamics of trust in the mentoring relationship.

3. Effective measure of success of collaborative learning.

Despite the wide acceptance of peer review in the research community, it has been highly criticized as encouraging bias in peer feedback. Many researchers have proposed approaches such as open, blind, and double-blind review approaches to enhance the peer review process. However, most of these researchers faulted only the peer review process with no definite conclusion on what constitutes a high-quality review and the most appropriate way to measure the success of the peer review process. Currently, the program chair, committee, or editorial team has the chance to evaluate the quality of reviews and may disqualify reviews or reviewers in extreme cases. However, this is a biased judgment because the authors are in the best position to judge the quality and helpfulness of the reviews. Therefore, we introduced back-evaluation of

reviews by the authors, as well as evaluation of the peer review process by all the participants, including authors and reviewers, as appropriate measures of success of the peer review process.

4. Identification of some factors that influence peer feedback in collaborative learning.

As a secondary contribution, we tested for some factors that can affect peer feedback in the modified peer review system. These are reciprocation of ratings, review lengths, and interpersonal trust score. Due to the fact that peers used pseudonyms, we were able to match the review ratings they gave their peers to the back-evaluation ratings they received from their peers. Therefore, we were able to check whether peers were reciprocating ratings by calculating the correlation between their review ratings and the corresponding back-evaluation ratings they gave their peers. Our findings showed that back-evaluation did not encourage reciprocation of ratings but rather encouraged participants to give honest and helpful feedback to their peer reviewers. Also, we were able to confirm that the quality of reviews cannot be evaluated by their length. In addition, we found that peers' rating behaviour was influenced by their interpersonal trust scores, and it also depended on their role in the relationship.

8.3 Implications of our Findings on Collaborative Learning

The findings of this dissertation have the following implications on the area of collaborative learning.

- 1 This research shows that in order to support group peer mentorship with peer-review systems, it is necessary to incorporate features that can allow participants to give both summative and formative feedback to their peers, as both mentors and mentees. Therefore, designers of collaborative learning tools should consider providing opportunity for peers to back-evaluate their peer feedback. Our studies found that back-evaluation encouraged peers to give thorough and helpful feedback to each other.
- 2 We found that grouping peers with different competence levels for peer review or mentorship will ensure that they get appropriate help. This finding reinforces the existing work on the advantages of forming heterogeneous groups for collaborative learning. However, it further shows that the designers of collaborative learning tools and instructors need to provide periodic evaluation of every group of learners to ensure that learning is taking place.

- 3 We found that the length of peer reviews is a weak predictor of the quality of reviews as perceived by the review recipients. We encourage conference and journal editors not to solely rely on the length of reviews to judge the quality of peer reviews.
- 4 We found that peers' rating behavior is influenced by their interpersonal trust score and it depends on the roles that they assume in the mentoring relationship. That is, the more trusting peers (mentees) have a high expectation of their peer mentors, therefore, they are more critical of their peer mentors when their expectations of their feedback are not met. Also, a highly trusting peer mentor will most likely offer high ratings to their mentees. Although the findings here still require more research, they can be useful in explaining the differences in the feedback that peers give to each other in mentorship and collaborative learning.
- 5 Our findings confirmed the advantages of anonymous review as it encourages peers in the mentor role to give thorough, honest and unbiased reviews. Also, it shows that the use of pseudonyms helps peers to accrue their peer feedback (ratings), which constitute their reputation among their peers, to be used in the future by the community of peers. Thus, designers of collaborative learning tools could leverage on the use pseudonyms to ensure learners are anonymous and are also able to retain their peer feedback in their pseudonyms for future decision relating to their reputation in the community of peers. Also in the research community, conference chairs and journal editors could benefit from accrued peers' ratings to determine the reputation of their peer reviewers when assigning papers for reviews.
- 6 Although our participants mentioned in their survey response that they preferred the peer review system to the wiki, we found that wiki writing exercises are actually helpful in mentoring students to improve their writing skill. However, we found that students require extrinsic motivation in form of feedback from their peers and an authority figure (e.g. instructor, TA) to enhance their quality of contribution to wiki. Therefore, designers of wikis for collaborative learning should not rely only on users' report when measuring their cognitive and affective states from using wikis. Also, designers and teachers should consider the use of appropriate motivational strategies while deploying wikis.

8.4 Limitations and Future Work

This research is a first effort towards establishing a framework for group peer mentorship using the modified peer review system. Our findings from the studies conducted have opened up opportunities for further research.

Although this research centred on mentorship, it provides a further improvement on collaborative learning, especially the peer review process. Fulfilling summative and formative objectives is vital in a collaborative learning venture that involves peers giving feedback on each other's work, irrespective of their demographics—age, gender, and cultural background, to mention but a few. Our findings showed that the mentorship model achieved these objectives in peers. However, in one of the studies with a multicultural class, we found that peers were mostly altruistic and that this encouraged them to give helpful and thorough feedback to their peers. We mentioned in section 4.3.4 that the results could have been affected by the collectivist culture of the participants (Hofstede, 2010). However, we suggested a further experiment to confirm this claim. We tried to repeat the experiment in another multicultural class in 2016, but the final survey participation was very low, and we could not obtain substantial data to support the reason for the difference in our findings. Therefore, we suggest in future work that the modified peer review system be used in a multicultural setting to measure the impact of peers' cultural backgrounds on their motivating factors.

In Chapter 7, we presented a study using a modified Wiki system to support group peer mentorship. Although our participants mentioned that they preferred the peer review system to the Wiki system, an implicit analysis of their grades from the use of the Wiki system showed an upward progression in their marks, which is an indication that learning actually took place. Even in the small study, we found that Wiki users required differentiated features to motivate participation, depending on the class of editors to which they belonged. In the future, we would like to try the use of a Wiki system in a large class broken into smaller subgroups (perhaps the size of 4, as in the peer-review group suggested). With this, we hope to test learning from the use of the Wiki in comparison with peer-review similar peer-group size, and also confirm whether our differentiated motivation features would encourage participants to contribute more and possibly change their opinions about the Wiki system.

In the six studies, we observed success in the implementation of the modified peer review process. In addition, we saw interesting results from the relationship between review and back-evaluation ratings, and also between peers' interpersonal trust scores and their back-evaluation ratings. We would like to extend this study to massive open online courses (MOOCs). Although we were able to evaluate our grouping algorithm with system-generated data from uniform, normal, and Pareto distributions, we would like to extend this algorithm to MOOCs with the hope of obtaining huge datasets from a large class. Research has shown that completion rate is low for MOOCs, and one reason is the lack of motivation on the part of learners. Although some other researchers have mentioned that peer evaluation would not necessarily improve participation and completion rates (Jordan, 2013), the modified peer review process contains back-evaluation by authors and overall evaluation by authors and reviewers that are not present in the peer evaluation strategy used in MOOCs. Also, we believe that the group-matching strategy would make some difference in the learning outcome, unlike the random approach used in most MOOCs at the moment. We had a preliminary result on the dynamics of trust in group peer mentorship. We would like to extend this study to MOOCs to test the significance of our preliminary findings.

8.5 Conclusion

In conclusion, this dissertation contributes to the area of collaborative learning, specifically to group formation, learners' motivation to participate and the measures of peers' learning from the collaboration. At the outset of this study, we observed that group peer mentorship is a relatively less explored area in collaborative learning. Hence, the initial challenge in finding a suitable tool to support our mentorship model. Over the six exploratory studies, we have observed that peer-review, with some modifications, has proven to be a good tool to organize, motivate and facilitate peer mentoring. However, the study also showed that the appropriate use of wiki as a tool to support peer-mentoring holds a promise, but needs further research. On the overall, this research area seems to be exciting since there is a complex interplay of factors, e.g. trust, reputation, culture and mechanism and system design at play, which deserve careful exploration and present opportunities for further research.

MY PEER-REVIEWED PUBLICATIONS WITH CONTENTS FROM THIS DISSERTATION

1. **Adewoyin, O.**; Phillips, F. and Vassileva, J. (2016). Dynamics of trust in group peer mentorship. First International Workshop on Intelligent Mentoring Systems (IMS 2016), held in conjunction with 13th International Conference on Intelligent Tutoring Systems (ITS2016) in Zagreb, Croatia, 9 pages.
2. **Adewoyin, O.**; Araya, R. and Vassileva, J. (2016). Peer Review in Mentorship: Perception of the Helpfulness of Review and Reciprocal Ratings. *Intelligent Tutoring Systems, Volume 9684, pp. 286-293.*
3. **Adewoyin, O.**; Wu, K. and Vassileva, J. (2015). Exploiting the use of Wikis to support collaborative learning: a case study of an undergraduate Computer Science class. 21st International Conference on Collaboration and Technology, N. Baloian et al. (Eds.): CRIWG 2015, LNCS 9334, *pp. 111–123.*
4. **Adewoyin, O.**; Wu, K. and Vassileva, J. (2015). Mapping Wiki User Contribution Types to Motivations for Participation: a Case Study. *Proc. Computer Supported Collaborative Learning CSCL'2015, Gothenburg, Sweden, pp. 649-650.*
5. **Adewoyin, O.** And Vassileva, J. (2014). Ethics of scientific peer review: Are we judging or helping the review recipients? 2014 IEEE International Symposium on Ethics in Engineering, Science, and Technology, Ethics'2014, *pp. 1-14.*
6. **Adewoyin, O.** (2014). Supporting Group Peer Mentorship System Using Online Peer Review. Doctoral Consortium: ACM Richard Tapia Celebration of Diversity in Computing, February 5-8, 2014, Seattle, WA, USA – 4 pages
7. **Adewoyin, O.** and Vassileva, J. (2013). Can Online Peer-Review Systems Support Group Mentorship? 16th International Conference on Artificial Intelligence in Education, Memphis 2013.
8. **Adewoyin, O.** and Vassileva, J. (2012). Recommendation, Trust and Reputation Management in Group Online Mentorship System. *Proc. PALE'2012 Workshop on*

Personalized and Adaptive Learning Environments, with UMAP'2012, Montreal, June 16, 2012, pp. 53-58 (CEUR Workshop Proceedings Vol. 872).

REFERENCES

1. Adamchik, V. (2005). Graph Theory. URL: https://www.cs.cmu.edu/~adamchik/21-127/lectures/graphs_5_print.pdf [Accessed: 30-July-2014].
2. Adewoyin, O. and Vassileva, J. (2013). Can Online Peer-Review Systems Support Group Mentorship? To appear in the proceedings of The 16th International Conference on Artificial Intelligence in Education (AIED 2013), Memphis, July 9-13, 2013, Memphis, Tennessee, USA.
3. Adewoyin, O. And Vassileva, J. (2014). Ethics of scientific peer review: Are we judging or helping the review recipients? 2014 IEEE International Symposium on Ethics in Engineering, Science, and Technology, Ethics'2014
4. Adewoyin, O.; Araya, R. and Vassileva, J. (2016). Peer review in mentorship: perception of the helpfulness of review and reciprocal ratings. 13th International conference on Intelligent Tutoring Systems, ITS 2016, 6-10 June, Croatia.
5. Adler, B. T.; Alfaro, L. De; Pye, I. and Raman, V. (2008). Measuring author contributions to the wikipedia. Proceedings of WikiSymp, ACM 2008, 10 pp.
6. Allen, S. J. (2007). Adult Learning Theory and Leadership Development. Kravis Leadership Institute, Leadership Review, Volume 7, Spring 2007, pp. 26-37.
7. Ally, M. (2008). Foundations of Educational Theory for Online Learning. In *The Theory and Practice of Online Learning*, Terry Anderson, Ed., May 2008.
8. Antin, J. And Cheshire, C. (2010). Readers are not free-riders: reading as a form of participation on Wikipedia. Proceedings of the 2010 ACM conference on computer supported cooperative work, pp. 127-130.
9. Antin, J.; Cheshire, C. and Nov, O. (2012). Technology-mediated contributions: Editing behaviors among new Wikipedians. Proceedings of the 2012 conference on computer supported cooperative work *CSCW 2012*, pp. 373-382.

10. Ardaiz-Villanueva, O.; Nicuesa-Chacon, X.; Brene-Artazcoz, O. (2011). Evaluation of computer tools for idea generation and team formation in project-based learning. *Journal of computers and education*, volume 56, pp. 700-711.
11. Astin, A. W. (1996). Involvement in Learning Revisited: Lessons We Have Learned. *Journal of College Student Development*, Volume 37(2), pp. 123-134.
12. Azimita, M. (1998). Peer interaction and problem solving: when are two heads better than one? *Child development*, volume 59, pp. 87-96.
13. Baggs, J. G.; Broome, M. E.; Dougherty, M. C.; Freda, M. C. and Kearney, M. H. (2008). Blinding in peer review: the preferences of reviewers for nursing journals. *Journal of advanced nursing*, Volume 64(2), pp. 131-138.
14. Balmaceda, J. M.; Schiaffino, S. and Diaz-Pace, J. A. (2014). Using constraint satisfaction to aid group formation in CSCL. *Inteligencia artificial*, Vol. 17(53), pp. 35-45.
15. Bandura, A. (1977a). *Social learning theory*. Englewood Cliffs, N.J.: Prentice-Hall
16. Bandura, A. (1977b). Self-efficacy: Toward a unifying theory of behavior change. *Psychological Review*, Volume 84(2), pp. 191-215.
17. Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ, Prentice Hall.
18. Belbin, R. M. (1983). *Team roles at work*. Oxford: Butterworth-Heinemann.
19. Ben-Ner, A. and Halldorsson, F. (2007). *Measuring Trust: Which Measure Can Be Trusted?* Paper provided by the human resources and labor studies, University of Minnesota in its working papers with number 0207.
20. Benson, G.; Noesgaard, C. and Drummond-Young, M. (2001). Facilitating small group learning. In Rideout E., ed. *Transforming nursing education through problem-based learning*. Sudbury MA.: Jones & Bartlett, pp. 75-102.

21. Bicknell-Holmes, T. and Hoffman, P. S. (2000). Elicit, engage, experience, explore: Discovery learning in library instruction. *Reference Services Review*, vol. 28(4), pp. 313-322.
22. Bloomberg University (2014). Misconceptions about mentoring. URL: <http://departments.bloomu.edu/reslife/studentstandards/Mentor%20Program%20Documents/mISCONCEPTIONS1.pdf> [Accessed: 14-January-2014].
23. Bodapati, A. V. (2009). Recommendation Systems with Purchase Data. *Journal of Marketing Research*, Volume XLV, pp. 77-93.
24. Bordage, G. and Regehr, G. (2006). To blind or not to blind? What authors and reviewers prefer. *Journal of Medical Education*, Volume 40(9), pp. 832-839.
25. Brindley, J. E.; Walti, C. and Blaschke, L. M. (2009). Creating Effective Collaborative Learning Groups in an Online Environment.
26. Bruff, D. (2005). The Assignment Problem and the Hungarian Method. URL: http://www.math.harvard.edu/archive/20_spring_05/handouts/assignment_overheads.pdf [Accessed:30-July-2014].
27. Bruner, J.S. (1967). *On knowing: Essays for the left hand*. Cambridge, Mass: Harvard University Press.
28. Bryant, S. L.; Forte, A. And Bruckman, A. (2005). Becoming Wikipedia: transformation of participation in a collaborative online encyclopaedia. *Proceedings of the 2005 international ACM SIGGROUP conference on supporting group work*, pp. 1-10.
29. Bryk, A. and B. Schneider (2002), *Trust in Schools: A Core Resource for Improvement*, New York: Russell Sage Foundation.
30. Castronova, J. (2002). Discovery learning for the 21st century: what is it and how does it compare to traditional learning in effectiveness in the 21st century. *Literature Reviews, Action Research Exchange (ARE)*, Vol. 1(2),

31. Cavanagh, R. R. (1996). Formative and summative evaluation in the faculty peer review of teaching. *Innovative higher education*, vol. 20, no. 4.
32. Cavanagh, R., Ellis, M., Layton, R., and Ardis, M. (2004). Automating the Process of Assigning Students to Cooperative-Learning Teams. *2004 American Society for Engineering Education Annual Conference & Exposition*, June 20-23, 2004, Salt Lake City, Utah.
33. Cestone, C.M., Levine, R.E., Lane, D.R.: Peer assessment and evaluation in team-based learning. *New Directions Teaching Learn.* 116, 69–78 (2008)
34. Chesler, N. C. and Chesler, M. A. (2002). Gender-Informed Mentoring Strategies for Women Engineering Scholars: On Establishing a Caring Community. *Journal of engineering education*, pp. 49-55.
35. Cho, K. and Schunn, C. D. (2007). Scaffolded writing and rewriting in the discipline: A web-based reciprocal peer review system. *Computers and Education*, vol. 48(3), pp. 409-426.
36. Cho, K.; Schunn, C. D. and Charney, D. (2006). Commenting on Writing: Typology and Perceived Helpfulness of Comments from Novice Peer Reviewers and Subject Matter Experts. *Written Communication*, Volume 23 Number 3, pp. 260-294.
37. Cho, K.; Schunn, C. D. and Wilson, R. (2006). Validity and reliability of scaffolded peer assessment of writing from instructor and student perspectives. *Journal of Educational Psychology*, vol. 98(4), pp. 891-901.
38. Christodoulopoulos, C. E. and Papaniklaou, K. A. (2007). A group formation tool in a e-learning context. *IEEE international conference on tools with artificial intelligence*, pp. 117-123.
39. Chua, F.; You, K. and Chuc, Z. H. (2009). Design and Implementation of a Collaborative learning System to Support Software Engineering Tasks, In: Kong, S.C., Ogata, H., Arnseth, H.C., Chan, C.K.K., Hirashima, T., Klett, F., Lee, J.H.M., Liu, C.C., Looi, C.K., Milrad, M., Mitrovic, A., Nakabayashi, K., Wong, S.L., Yang, S.J.H. (eds.) (2009). *Proceedings of the 17th International Conference on Computers in Education*. Hong Kong: Asia-Pacific Society for Computers in Education.

40. Chubin, D. and Hackett, E. (1990). *Peerless Science: Peer Review and US Science Policy*. Albany: SUNY Press.
41. Clark, R.; Andrews, J. and Davies, K. (2011). *Peer Mentoring Works! Institutional Manual*. URL:http://www.heacademy.ac.uk/assets/documents/what-works-student-retention/Aston_Peer_Mentoring_Institutional_Manual.pdf [Accessed: 22nd-July-2014].
42. Cocea, M. and Magoulas, G. D. (2012). User behaviour-driven group formation through case-based reasoning and clustering. *Expert systems with applications*, vol. 39, pp. 8756-8768.
43. Cohen, N. (2014). *Wikipedia vs. the Small Screen*. The New York Times. URL: http://www.nytimes.com/2014/02/10/technology/wikipedia-vs-the-small-screen.html?_r=2 [Retrieved: 14 April 2014].
44. Cole, M. (2009). Using wiki technology to support students' engagement: Lessons from the trenches. *Journal of Computers and Education*, volume 52(2009), pp. 141-146.
45. Cook, J., & Wall, T. (1980). New work attitude measures of trust, organizational commitment, and personal need fulfillment. *Journal of Occupational Psychology*, 53, 39-52.
46. Craig, M.; Horton, D. and Pitt, F. (2010). Forming reasonably optimal groups (FROG). *GROUP 2010*, pp. 141-150.
47. Cross, A.; Bayyapunedi, M.; Ravindran, D.; Cutrell, E. and Thies, W. (2014). VidWiki: Enabling the crowd to improve the legibility of online educational videos. *Proceedings of the 2014 conference on computer supported cooperative work CSCW 2014*, pp. 1167-1175.
48. Crow, T. (2010). Learning no matter where you are: Q & A with Chris Dede. *Journal of Staff Development*. 31(1).
49. Csernica, J.; Hanyka, M.; Hyde, D.; Shooter, S.; Toole, M. and Vigeant, M. (2002). *Practical guide to teamwork*, version 1.1. College of Engineering, Bucknell University.
50. Davis, B. G. (1993). *Tools for Teaching*. Jossey-Bass Inc., San Francisco: California.

51. de Faria, E.S.J., Adan-Coello, J.M., and Yamanaka, K. (2006). Forming Groups for collaborative Learning in Introductory Computer Programming Courses Based on Students' Programming Styles: An Empirical Study. *Proceedings of the ASEE/IEEE Frontiers in Education Conference*, San Diego, CA, USA, 6-11.
52. Deal, A. (2009). A teaching with technology white paper: collaboration tools. URL: http://www.cmu.edu/teaching/technology/whitepapers/CollaborationTools_Jan09.pdf [Accessed: 18-December-2014].
53. DeVine, C. (2013). The skills both online students and teachers must have. URL: <http://www.edudemic.com/the-skills-both-online-students-and-teachers-must-have/> [Accessed: 06-04-2015].
54. Dirks, K. T. (1999). The effects of interpersonal trust on work group performance. *Journal of applied psychology*, volume 84, pp. 445-455.
55. Do, E. Y. (2003). Afterword- Why peer review journals? *International journal of architectural computing*, issue 02, volume 01, pp. 253-264.
56. Doucer, J. R. (2008). Paper Rating vs. Paper Ranking. In *Proceedings of 1st Workshop on Organizing Workshops, Conferences, and Symposia for Computer Systems (WOWCS)*.
57. Dow, S. P.; Kulkarni, A.; Klemmer, S. R. And Hatmann, B. (2012). Shepherding the crowd yields better work. In: *Proceedings of ACM CSCW12 Conference on Computer-Supported Cooperative Work 2012*. pp. 1013-1022.
58. EasyChair (2014). Overview of EasyChair. URL: <https://www.easychair.org/easychair.cgi> [Accessed: 24-July-2014].
59. EduTech Wiki (2008). Mentoring. URL: <http://edutechwiki.unige.ch/en/Mentoring> [Accessed: 15-December-2014].
60. Engel, C. E. (1991). Not just a method but a way of learning. In Boud D., Feletti G. P., eds. *The challenge of problem-based learning*. London: Kogan page, 1999, pp. 22-33.

61. Engelbrecht, E. (2003). A Look at E-Learning Models: Investigation their Value for Developing an E-Learning Strategy. *Progressio* 2003, Volume 25, Number 2, pp 38-47
62. Engstrom, M.E. and Jewett, D. (2005). Collaborative learning the wiki way. *TechTrends*, vol. 49 (6), pp. 12-15.
63. Ensher, E. A.; Heun, C. and Blanchard, A. (2003). Online Mentoring and Computer-mediated Communication: New Directions in Research. *Journal of Vocational Behavior*, Volume 63(2000), pp. 264-288.
64. Farrell, K.: Collegial feedback on teaching: a guide to peer review. Centre for the Study of Higher Education, The University of Melbourne (2011)
65. Fernandez, C. E. and Yu, J. (2007). Peer review of teaching. *The journal of chiropractic education*, vol. 21(2), pp. 154-161.
66. Frank, A. (2004). On Kuhn's Hungarian Methods - A tribute form. Egrervary Research Group (EGRES) Technical Report No.2004-14 URL: <http://www.cs.elte.hu/egres/tr/egres-04-14.pdf> [Accessed: 30-July-2014].
67. Gale, D. and Shapley, L. S. (1962). College admissions and the stability of marriage. *Ame Maths monthly*, volume 69, pp. 9-15.
68. Gartner, A.; Kohler, M. and Riessman, F. (1971) Children teach children. Learning by teaching. Harper & Row, New York 1971.
69. Gaved, M.; Heath, T. and Eisenstadt, M. (2006). Wikis of locality: insights from the open guides. Proceedings of the 2006 international symposium on wikis.
70. Gayle Morris Sweetland Centre for writing. Using peer review to improve student writing. University of Michigan. URL: <http://www.lsa.umich.edu/UMICH/sweetland/Home/For%20Students/Writing%20References%20and%20Resources/Using%20Peer%20Review%20to%20Improve%20Student%20Writing.pdf> [Accessed:22-December-2014].

71. Geiger, R. S. and Halfaker, A. (2013). Using edit sessions to measure participation in Wikipedia. Proceedings of the 2013 conference on computer supported cooperative work, pp. 861-870.
72. Geiger, R. S. And Ribes, D. (2010). The work of sustaining order in Wikipedia: the banning of a scandal. Proceedings of the 2010 conference on computer supported cooperative work *CSCW 2010*.
73. Gibson, M.; Spong, C. Y.; Martin, S. And Scott, J. R. (2008). Author perception of peer review. *Obstetrics and gynecology*, Volume 112(3), pp. 646-652.
74. Graf, S., and Bekele, R. (2006). Forming Heterogeneous Groups for Intelligent Collaborative Learning Systems with Ant Colony Optimization. *Intelligent tutoring systems*, vol. 4053, pp. 217-226.
75. Greene, W. A. (2001). Genetic algorithms for partitioning sets. *Journal on artificial intelligence tools*, volume 10(1-2), pp. 225-241.
76. Greer, J.; McCalla, G.; Collins, J.; Kumar, V.; Meagher, P. and Vassileva, J. (1998). Supporting Peer Help and Collaboration in Distributed Workplace Environments. *International Journal of Artificial Intelligence in Education*, Volume 9, pp. 159-177.
77. Grusec, J. E. (1992). Social learning theory and developmental psychology: The legacies of Robert Sears and Albert Bandura. *Developmental Psychology*, Volume 28, Number 5, pp. 776-786.
78. Gusfield, D. and Irving, R. W. (2003). The stable marriage problem: structure and algorithms. *Foundations of computing series*, MIT press. page 9.
79. Gutknecht-Gmeiner, M. (2005). Peer review in education. Peer review in initial VET, Leonardo da Vinci project, Austria.
80. Guzdial, M.; Ludovice, P.; Realff, M.; Morley, T.; Carroll, K. and Ladak, A. (2001). The challenge of collaborative learning in engineering and math. 31st Annual frontiers in education conference, vol. 1, pp. 24-29.

81. Guzdial, M.; Ludovice, Realff, M.; Morley, T. and Carroll, K. (2002). When collaboration doesn't work. Proceedings of the international conference on learning sciences, pp. 125-130.
82. Gyambrah, M. K. (2007) E-learning Technologies and its Application in Higher Education: A Descriptive Comparison of Germany, UK and USA. Inaugural PhD Dissertation. URL: http://edoc.ub.uni-muenchen.de/7358/1/Gyambrah_Martin_K.pdf [Accessed: 09-October-2012].
83. Hamalainen, H.; Tarkkonen, J.; Heikkinen, K.; Ikonen, J. and Porras, J. (2009). Use of peer review system for enhancing learning of programming. In the proceeding of the 9th IEEE international conference on advanced learning technologies, pp. 658-660.
84. Hanson, R. (2006). Collaborative learning and peer reviews in special education. Action research.
85. Hart-Davidson, W.; McLeod, M.; Klerkx, C. and Wojcik, M. (2010). A method for measuring helpfulness in online peer review. SIGDOC, 2010 - Proceedings of the 28th ACM International Conference on Design of Communication, pp. 115-121.
86. Hayes, J. R. and Flower, L. (1980). Identifying the organization of writing processes. In Gregg, L. and Steinberg, E. (Eds.) Cognitive processes in writing: an interdisciplinary approach, pp. 3-30.
87. Henry, T. R. (2013). Creating effective student groups: an introduction to groupformation.org. SIGCSE '13 Proceeding of the 44th ACM technical symposium on Computer science education, pp. 645-650.
88. Herrera, C.; Vang, Z. and Gale, L. Y. (2002). Group Mentoring: A study of Mentoring Groups in Three Programs. Prepared for The National Mentoring Partnership's Public Policy Council. Page17.
89. Herrington, J. and Oliver, R. (1995). Critical characteristics of situated learning: Implications for the instructional design of multimedia. In: ASCILITE 1995 Conference, pp. 253-262.
90. Hirschauer, S. (2010). Editorial judgements: A praxeology of 'voting' in peer review. *Social Studies of Science*, vol. **40** (1),pp. 71–103.

91. Ho, A. D.; Reich, J.; Nesterko, S.; Seaton, D. T.; Mullaney, T.; Waldo, J. and Chuang, I. (2014). HarvardX and MITx: The first year of open online courses (HarvardX and MITx Working Paper No. 1).
92. Ho, T.; Shyu, S. J.; Wang, F. and Li, C. T. (2009). Composing High-heterogeneous and High-interaction groups in collaborative learning with particle swarm optimization. World congress on computer sciences and information engineering, pp. 607-611
93. Hodges, B. (2009). Factors that can influence mentorship relationships. Paediatric nursing, Volume 21 (6), pp. 32-35.
94. Hofstede, G.; Hofstede, G. J. and Minkov, . (2010). Cultures and Organizations: Software of the Mind. Revised and expanded 3rd Edition. 550 pages. New York: McGraw-Hill USA.
95. Hoisl, B.; Aigner, W. and Miksch, S. (2007). Social Rewarding in Wiki Systems – Motivating the Community. Proceedings of the 2nd international conference on online communities and social computing, pp. 362-371.
96. Hoppe, H. U. (1995). Use of multiple student modeling to parameterize group learning. Proceedings of the 7th world conference on artificial intelligence in education (AIED).
97. Houry, D.; Green, S. and Callaham, M. (2012). Does mentoring new peer reviewers improve review quality? A randomized trial. BMC Med Educ, vol. 12(83).
98. Hubscher, R. (2010). Assigning students to groups using general and context-specific criteria. IEEE transactions on learning technologies, vol. 3, no 3, pp. 178-189.
99. Hutchings, P. (1994): Peer Review of Teaching. "From Idea to Prototype", in: AAHE Bulletin, November 1994.
100. IJSER (2016). Peer Review Publication Process. International Journal of Scientific and Engineering Research. URL: <http://www.ijser.org/peer-review-journal.aspx> [Accessed: 05-March-2106]

101. Ikeda, M., Go, S., & Mizoguchi, R. (1997). Opportunistic group formation. *Proceedings of Artificial Intelligence in Education (AIED): Knowledge and Media in Learning Systems*, Amsterdam: IOS, 166-174.
102. Illes, J.; Lau, P. and Giacino, J. T. (2008). Brain Imaging: Understanding the Basics. The University of British Columbia. URL: <http://www.acrm.org/pdf/BrainImagingFAQ.pdf> [Accessed: 11-February-2013].
103. Inaba, A., Supnithi, T., Ikeda, M., Mizoguchi, R., and Toyoda, J. (2000). How Can We Form Effective Collaborative Learning Groups? *Proceedings of the 5th International Conference on Intelligent Tutoring Systems*, 282-291
104. Inaba, A.; Tamura, T.; Ohkubo, R.; Ikeda, M. and Mizoguchi, R. (2001). Design and analysis of learners' interaction based on collaborative learning ontology. The European conference on computer-supported collaborative learning, pp. 308-315.
105. Interpersonal trust scale. URL: <http://higher.mheducation.com/sites/dl/free/0072563974/87095/ch07.html> [Accessed: 01-September-2015]
106. Iwama, K. and Miyazaki, S. (2008). A survey of the stable marriage problem and its variants. International conference on informatics education and research for knowledge-circulating society, pp. 131-136.
107. Jordan, K. (2013). MOOC Completion Rate: The Data. URL: <http://www.katyjordan.com/MOOCproject.html> [Accessed: 06-04-2015].
108. Joyce, E., & Kraut, R. E. (2006). Predicting continued participation in newsgroups. *Journal of Computer-Mediated Communications*, volume 11, pp. 723 - 747.
109. Justice, A. C.; Cho, M. K.; Winker, M. A.; Berlin, J. A.; Rennie, D. And Peer Investigators (1998). Does masking author's identity improve the quality of peer review? A randomised controlled trial. *The journal of American medical association (JAMA)*, Volume 280, number 3, pp. 240-242

110. Kinsey, J.; Carleo, J.; O'Neill, B. and Polanin, N. (2010). The Wiki as a Time-Saving Mentoring Tool. *Journal of Extension*, volume 48, number 2, pp. 1-4.
111. Knox, J. D. E. (1980). How to use modified essay questions. *Med Teach*, Vol. 2, pp. 20-24.
112. Kollock, P. (1999). The economies of online cooperation: gifts and public goods in the cyberspace, In: Smith, Marc and Kollock, Peter. *Communities in cyberspace*. London, Routledge, *pp.* 220-239.
113. Komiskey, L.; Brinkerhoff, D.; Miller, H.; Morrison, J. and Morrison, M. (2010). A Peer Review System to Enhance Collaborative Learning: Testing and Preliminary Evaluation. 2010 Midwest instruction and computing symposium.
114. Konert, J.; Burlak, D. and Steinmetz, R. (2014). The group formation problem: an algorithmic approach to learning group formation. *European conference on technology enhanced learning, EC-TEL, LNCS 8719*, *pp.* 221-234.
115. Kop, R. and Hill, A. (2008). Connectivism: Learning theory of the future or vestige of the past? In "The International Review of Research in Open and Distance Learning, Vol 9, No 3 (2008).
116. Kram, K. E. and Isabella, L. A. (1985). Mentoring alternatives: The role of peer relationships in career development. *Academy of management Journal*, *Volume* 28, pp. 110–132.
117. Kronick, D. A. (1990). Peer review in 18th century scientific journalism. *Journal of American medical association (JAMA)*, volume 263, pp. 1321-1322.
118. Krzyzanowski, P. (2014). *Operating Systems*. URL: <https://www.cs.rutgers.edu/~pxk/416/notes/07-scheduling.html> [Accessed: 31-July-2014].
119. Kuhn, D. (1972). Mechanisms of change in the development of cognitive structures. *Child development*, volume 43, *pp.* 833-844.

120. Kuhn, H. W. (1955). The Hungarian Method for the assignment problem. *Naval Research Logistics Quarterly*, Volume 2, pp. 83–97.
121. Kuhn, H. W. (1956). Variants of the Hungarian method for assignment problems. *Naval Research Logistics Quarterly*, Volume 3, pp. 253–258.
122. Kumar, V. S. (1996), Computer-Supported Collaborative Learning: Issues for Research, Department of Computer Science, University of Saskatchewan. URL: <http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=2B4B8E311DFE4800712DF8BBE064F96E?doi=10.1.1.150.6693&rep=rep1&type=pdf> [Accessed: 09-July-2014].
123. Lan, Y.; Wang, J.; Hsu, S. and Chan, T. (2011). Peer feedback in online writing system. In: M. Chang et al. (Eds.): *Edutainment 2011*, LNCS 6872, pp. 126–129, 2011.
124. Lateef, F. (2010). Simulation-based learning: just like the real thing. *Journal of emerg trauma shock*, vol. 3(4), pp. 348-352.
125. Lave, J. and Wenger, E. (1991) *Situated Learning. Legitimate peripheral participation*, Cambridge: University of Cambridge Press.
126. Lawrence, E. C.; Levy, M.; Martin, N.; and Strother-Taylor, J. (2008). *Case Studies in Youth Mentoring One-on-One and Group Mentoring : An Integrated Approach*. U.S. Department of Education: Office of Safe and Drug-Free Schools. Mentoring Resource Center.
127. Leck, J. D. and Wood, P. M. (2013). Forming trust in e-mentoring: a research agenda. *American Journal of Industrial and Business Management*, 2013, Volume 3, pp.101-109
128. Leung, K. and Chu, SKW. (2009). Using Wikis for collaborative learning: a Case study of an undergraduate students' group project in Hong Kong. *The 6th International Conference on Knowledge Management (ICKM 2009)*, Hong Kong. In *Proceedings of ICKM, 2009*, pp. 1-13.
129. Levine, R.E.: *Peer evaluation in team-based learning* (2010). https://training.health.ufl.edu/handouts/FacDev/TBL_Chapter9.pdf. (Accessed on 17 October 2015)

130. Li, C.; Dong, Z. J.; Untch, R. and Chasteen, M. (2010). Facilitating peer review in an online collaborative learning environment for computer science students. *International conference on frontiers in education: computer science and computer engineering*.
131. Ling, K.; Beenen, G.; Ludford, P. J.; Wang, X.; Chang, K.; Li, X.; Cosley, D.; Frankowski, D.; Terveen, L.; Rashid, A.; Resnick, P. and Kraut, R. (2005). Using social psychology to motivate contributions to online communities. *Journal of computer mediated communication*, volume 10(4).
132. Lipponen, L. (2002). Exploring Foundations for Computer-Supported Collaborative Learning. *Proceedings of the Conference on Computer Support for Collaborative Learning: Foundations for a CSCL Community*, pp. 72-81.
133. Liu, Y.; Liu, Q.; Wu, R.; Chen, E.; Su, Y.; Chen, Z. and Hu, G. (2016). Collaborative learning team formation: a cognitive modeling perspective. *Database Systems for Advanced Applications*, Vol. 9643, pp. 383-400.
134. Lo, S.; Phan, R. C. W. and Goi, B. (2007). On the security of a popular web submission and review software for cryptography conferences. *Workshop on Information security applications* URL: <http://lasec.epfl.ch/pub/lasec/doc/LPG07.pdf> [Accessed: 24-July-2014]
135. Logan, D. (2001). E-learning in the knowledge age. *Gartner Symposium Itxpo 2001*. 30 July –1 August 2001, Johannesburg, South Africa.
136. Long, M.; Wood, C.; Littleton, K.; Passenger, T. and Sheeh, K. (2011). *The Psychology of Education*. Routledge, 2nd Edition.
137. Lugano, G., Nokelainen, P., Miettinen, M., Kurhila, J., and Tirri, H. (2004). On the Relationship between Learners' Orientations and Activity in CSCL. *Proceedings of the IEEE International Conference on Advanced Learning Technologies (ICALT'04)*, LosAlamitos: IEEE Computer Society Press, 759-761.
138. Lundstrom, K., & Baker, W. (2009). To give is better than to receive: The benefits of peer review to the reviewer's own writing. *Journal of Second Language Writing*, 18(1), 30-43.

139. Lunsford, A. (1991). "Collaboration, control and the idea of a writing center". *The writing center journal*, volume 12(1), pp. 3-10.
140. MacDonald, C. J.; Stodel, E. J.; Farres, L. G.; Breithaupt, K. and Gabriel, M. A. (2001). The demand-driven learning model: a framework for web-based learning. *The Internet and Higher Education Volume 4*(2001):pp 9-30.
141. Mackay, T. (2007). The social informatics of blog and wiki communities: authoring communities of practice (CoPs). *Proceedings of Canadian Association for Information Science*. Montreal Canada, pp. 1-14.
142. Madden, N. A. and Slavin, R. E. (1983). Effects of cooperative learning on the social acceptance of mainstreamed academically handicapped students. *The journal of special education*, volume 7(2), pp. 171-182.
143. Majchrzak, A.; Wagner, C. and Yates, D. (2006). Corporate wiki users: results of a survey. *Proceedings of the 2006 international symposium on wikis*.
144. Manske, S.; Hecking, T.; Chounta, I.; Werneburg, S. and Hoppe, H. U. (2015). Using differences to make a difference: a study on heterogeneity of learning groups. *11th International Conference on Computer Supported Collaborative Learning (CSCL 2015)*, vol. 1, no. 2, pp. 182-189.
145. Masthoff, J. (2002). Modeling a Group of Television Viewers. In *Proceedings of the Future TV: Adaptive Instruction in your Living Room Workshop*, associated with ITS02, pp. 34-42.
146. Masthoff, J. (2004). Group modeling: Selecting a sequence of television items to suit a group of viewers. *User Modeling and User Adapted Interaction*, Volume 14, pp. 37-85. © Springer-Verlag
147. Masthoff, J. (2010). Group recommender systems: Combining individual models. In Ricci, F., Rokach, L., Shapira, B., Kantor, P. (Eds.), *Recommender Systems Handbook* (pp. 677-702). Dordrecht, Netherlands: Springer Science+Business Media.

148. Masthoff, J. and Gatt, A. (2006). In Pursuit of Satisfaction and the Prevention of Embarrassment: Affective State in Group Recommender Systems. *Journal of User Modeling and User-Adapted Interaction*, Volume 16, Issue 3-4, pp. 281-319
149. Mayer, R. C.; Davis, J. H. and Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, volume 20, pp. 709-734.
150. McAllister, D. (1995). Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of Management Journal*, 38, 24-59.
151. McCarthy, J. and Anagnost, T. (1998). MusicFX: An Arbiter of Group Preferences for Computer Supported Collaborative Workouts. *CSCW*, Seattle, WA. (1998), pp. 363-372.
152. McEvily, B., V. Perrone, and A. Zaheer (2003), "Trust as an organizing principle", *Organization Science*, volume 14(1): pp. 91-103.
153. McLean, I., and F. Hewitt (eds.) (1994). *Condorcet: Foundations of Social Choice*
154. McLehan, H. (1991). Virtual environments and situated learning. *Multimedia Review*, vol. 2, No. 3, pp. 30-37.
155. McNutt, R. A.; Evans, A. T.; Fletcher, R. H. and Fletcher, S. W. (1990). The Effects of Blinding on the Quality of Peer Review: A Randomized Trial. *Journal of American medical association (JAMA)*, volume 263(10), pp. 1371-1376.
156. Menditto, A.; Patriarca, M. and Magnusson, B. (2007). Understanding the meaning of accuracy, trueness and precision. *Discussion Forum, Accreditation and Quality Assurance*, Volume 12, Issue 1, pp 45-47.
157. Minor, F. D. (2007). Building effective peer mentor programs. URL: <http://evergreen.edu/washingtoncenter/docs/monographs/lcsa/lcsa4building.pdf> [Accessed: 17-February-2015].
158. Mishra, A. K. (1993). Breaking down organizational boundaries during crisis: The role of mutual trust. Paper presented at the 1993 Academy of Management Meeting, Atlanta, GA.

159. Moreno, J.; Ovalle, D. A. and Vicari, R. M. (2012). A genetic algorithm approach for group formation in collaborative learning considering multiple student characteristics. *Computers and Education*, vol. 58, pp. 560-569.
160. Myers, K. A. and Distlehorst, L. (2000). Groups in problem-based learning: essential elements in theory and practice. In Evensen D., Hmelo C., eds. *Problem based learning: a research perspective on learning interactions*, Mahwah, NJ.: Lawrence Erlbaum, pp. 167-184.
161. Newman, M. J. (2005). Problem based learning: an introduction and overview of the key features of the approach. *Journal of veterinary medical education, JVME*, Vol. 32(1), pp. 12-20.
162. Nilson, L. B. (2003). Improving student peer feedback. *College teaching*, vol. 51:1, pp. 34-38
163. Nonnecke, B. and Preece, J. (2000). Lurker demographics: counting the silent. In *CHI 2000 Conference Proceedings. Conference on Human Factors in Computing Systems. CHI2000. The Future is Here. The Hague, Netherlands: ACM (2000) pp.73-80*
164. Noorian, Z.; Marsh, S. and Fleming, M. (2011). Multi-Layer Cognitive Filtering by Behavioral Modelling. *Proc. of 10th International conference on autonomous agents and multiagent systems (AAMAS 2011)*, pp. 871-878.
165. O' Conner, M.; Cosley, D.; Konstan, J.A. and Riedl, J. (2001). PolyLens: A Recommender System for Groups of Users. *Proceedings of the 7th Conference on European Conference on Computer Supported Cooperative Work*, pp. 199-218.
166. Ormrod, J.E. (1999). *Human learning*, 3rd edition. Upper Saddle River, NJ: Prentice-Hall
167. Osmond, D. H. (1983). Malice's wonderland: research funding and peer review. *Journal of neurobiology*, vol. 14 no. 2, pp. 95-112.
168. Ounnas, A., Davis, H. C., and Millard, D. E. (2009). A Framework for Semantic Group Formation in Education. *Educational Technology & Society*, 12 (4), pp. 43–55.

169. Ounnas, A.; Davis, H. C. and Millard, D. E. (2007). Towards Semantic Group Formation. IEEE international conference on advanced learning technologies (ICALT), pp. 825-827.
170. Ounnas, A.; Davis, H. C. and Millard, D. E. (2009). A framework for semantic group formation in education. Educational technology and society, vol. 12(4), pp. 43-55.
171. Ozdogan, C. (2011). Round robin scheduling. URL: <http://siber.cankaya.edu.tr/OperatingSystems/ceng328/node125.html> [Accessed: 31-July-2014].
172. Painvin, C.; Neufeld, V. R. and Norman, G. (1979). The triple jump exercise: a structured measure of problem-solving and self-directed learning. In proceedings of the 18th conference of research in medical education. Association of American medical colleges, pp. 73-77.
173. Palloff, R. M. and Pratt, K. (2005). Learning together in community: Collaboration online. 20th Annual conference on distance teaching and learning.
174. Panther Learning (2014). Why SWoRD peer assessment? URL: <https://sword.pantherlearning.com/wordpress/why-sword-peer-review/> [Accessed: 26-July-2014].
175. Pearce, J., Mulder, R., Baik, C.: Involving students in peer review: Case studies and practical strategies for University teaching. University of Melbourne, Victoria (2009). http://www.cshe.unimelb.edu.au/resources_teach/teaching_in_practice/docs/Student_Peer_Review.pdf. (Accessed on 17 October 2015)
176. Peerceptiv (2016). URL: <http://www.peerceptiv.com/wordpress/about-us/> [Accessed: 14-March-2016].
177. Peltola, M. (2014). Hungarian method: a cost matrix. URL: http://www.ee.oulu.fi/~mpa/matreng/ematr1_2.htm [Accessed: 30-July-2014].
178. Peters D. and Ceci S. (1982). Peer-review practices of psychological journals: the fate of submitted articles, submitted again. Behav Brain Sci 1982;vol. 5, pp.187-255

179. Piaget, J. (1928). *The judgement and reasoning in children*. London: Routledge and Kegan.
180. Pini, M. S.; Rossi, F.; Venable, K. B. and Walsh, T. (2009). Manipulation and gender neutrality in stable marriage procedures. *Proc. of 8th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2009)*, Decker, Sichman, Sierra and Castelfranchi (eds.), *pp.*
181. Pizzato, L.; Rej, T.; Chung, T.; Koprinska, I. and Kay, J. (2010). RECON: A Reciprocal Recommender for Online Dating. Xavier Amatriain, Marc Torrens, Paul Resnick, Markus Zanker (Eds.): *Proceedings of the 2010 ACM Conference on Recommender Systems, RecSys 2010*, pp. 207-214.
182. Pollock, P. H.; Hamann, K. and Wilson, B. M. (2009). Comparing the benefits of small-group and large-class discussions. *APSA 2009 Toronto Meeting Paper*. McKisack, C. And Waller, G. (1997). Factors influencing the outcome of group psychotherapy for bulimia nervosa. *International Journal of Eating Disorder*, Volume 22(1): pp. 1-13.
183. Pompper, D. and Adams, J. (2006). Under the microscope: Gender and mentor-protégé relationships. *Public Relations Review* (Science Direct), Volume 32 (32):pp. 309–315.
184. Pretz, K. (2014). Low completion rate for MOOCs. URL: <http://theinstitute.ieee.org/ieee-roundup/opinions/ieee-roundup/low-completion-rates-for-moocs> [Accessed: 06-04-2015].
185. Quadackers, L.; Groot, T. and Wright, A. (2014). Auditors' professional skepticism: Neutrality versus presumptive doubt. *Contemporary accounting research*, vol. 31(3), pp. 639-657.
186. Rafaeli, S. and Ariel, Y. (2008). Online motivational factors: Incentives for participation and contribution in Wikipedia. In Barak, A. *Psychological aspects of cyberspace: Theory, research, applications*. Cambridge, UK: Cambridge University Press. ISBN 0-521-87301-0.
187. Ragins, B. R. and Kram, E. D. (2007). *The handbook of mentoring at work: theory, research and practice*. Thousand Oaks, CA: Sage.
188. Ralph, E. and Walker, K. (2011). *Mentorship across Professions*. *Proceedings of the 1st Forum on Mentorship in Professions, Saskatchewan 2011*.

189. Raman, M.; Ryan, T. and Olfman, L. (2005). Designing knowledge management systems for teaching and learning with wiki technology. *Journal of Information systems education*, volume 16(3), pp. 311-320.
190. Ranalli, B. (2011). A Prehistory of Peer Review: Religious Blueprints from the Hartlib Circle. *Spontaneous Generations: A Journal for the History and Philosophy of Science*, Vol. 5, No. 1 (2011) pp. 12-18.
191. Redmond, M.A. (2001). A computer program to aid assignment of student project groups. *Proceedings of the 32nd SIGCSE Technical Symposium on Computer Science Education*, Charlotte, NC, USA.
192. Reich, J. (2014). MOOC Completion and Retention in the Context of Student Intent. URL: <http://www.educause.edu/ero/article/mooc-completion-and-retention-context-student-intent> [Accessed:06-04-2015].
193. Rogoff, B. (1991). Social interaction as an apprenticeship in thinking: Guided participation in spatial planning. In L. Resnick and J. Levine and S. Teasley. *Perspectives in socially shared cognition*, pp. 349-364.
194. Ross, C. M. (2014). Recreational sports programming. URL: <http://www.indiana.edu/~r324/roundrobinscheduling.pdf>[Accessed: 31-July-2014].
195. Rotter, J. B. (1980) Interpersonal trust, trustworthiness and gullibility. *American psychologist*, volume 35(1), pp. 1-7.
196. Rowe, M.; Fernandez, M. and Alani, H. (2013). Modelling and analysis of user behavior in online communities. *STC Social Networking E-Letter Vol.1 No.2* URL: <http://stcsn.ieee.net/e-letter/vol-1-no-2> [Accessed: 27-June-2014].
197. Ryvkin, D. (2005). The Predictive Power of Noisy Elimination Tournaments. *CERGE-EI, WORKING PAPER SERIES (ISSN 1211-3298)*. URL: <https://www.cerge-ei.cz/pdf/wp/Wp252.pdf> [Accessed: 06-August-2014].
198. Sabater, J. and Sierra, C. (2005). Review on computational trust and reputation models. *Artificial intelligence review*, vol. 24(1), pp. 33-60.

199. Sachs, J., Parsell, M. (Eds.). Peer review of learning and teaching in higher education: international perspectives. Berlin, Germany: Springer (2014).
200. Sadeghi, H. and Kardan, A. A. (2016). Toward effective group formation in computer-supported collaborative learning. *Interactive learning environments*, Vol. 24, no. 3, pp. 382-395.
201. Sammons, A. (2009). Approaches to Psychology. URL: <http://www.psychotron.org.uk/newResources/perspBiological.html> [05-January-2013].
202. Sarkar, A.; Seth, D.; Basu, K. and Acharya, A. (2015). A new approach to collaborative group formation. *International journal of computer applications*, Vol. 128, No. 3.
203. Savin-Baden, M. (2000). Problem-based learning in higher education: untold stories. Society for research in higher education, open university press, 2000.
204. Schank, R. and Cleary, C. (1995). Engines for education. Lawrence Erlbaum Associates, Hillsdale, NJ
205. Schmidt, H. and Moust, J. (2000). Factors affecting small-group tutorial learning: a review of research. In Evensen D., Hmelo C., eds. *Problem based learning: a research perspective on learning interactions*, Mahwah, NJ.: Lawrence Erlbaum, pp. 19-51.
206. Schroter, S., Black, N., Evans, S., Carpenter, J., Godlee, F., Smith, R.: Effects of training on quality of peer review: randomised controlled trial. *BMJ* 2004(328), 673 (2004)
207. Schunk, D. H. (2008). *Learning theories: an educational perspective*, 5th Edition. Upper Saddle River, N.J.; London: Merrill Prentice Hall; Chapters 4 & 5.
208. Searby, M., Ewers, T.: An evaluation of the use of peer assessment in higher education: a case study in the school of music. *Assess. Eval. Higher Educ.* 22(4), 371–383 (1997). Kingston University
209. Sears, R. (1951). A Theoretical Framework for Personality and Social Behavior. *American Psychologist*, Volume 6, pp. 476–483.

210. Siemens, G. (2004). Connectivism: a learning theory for the digital age. URL: http://www.ingedewaard.net/papers/connectivism/2005_siemens_ALearningTheoryForTheDigitalAge.pdf [Accessed: 18-December-2014].
211. Smith, R. (2006). Peer review: a flawed process at the heart of science and journals. *Journal of the Royal Society of Medicine*, Volume 99, pp. 178-182.
212. Soh, L. K. (2004). On Cooperative Learning Teams for Multiagent Team Formation. *Technical Report WS-04-06 of the AAAI'2004 Workshop on Forming and Maintaining Coalitions and Teams in Adaptive Multiagent Systems*, July 26, San Jose, CA, pp. 37-44.
213. Soh, L. K., Khandaker, N., Liu, X., and Jiang, H. (2006). A Computer-Supported Cooperative Learning System with MultiagentIntelligence. *Proceedings of the 5th international joint conference on Autonomous agents and multiagent systems*, New York:ACM, pp. 1556-1563.
214. Sorensen, J. (2008). Measuring Emotions in a Consumer decision-making Context – Approaching or Avoiding. Working Paper Series, Department of Business Studies, Aalborg University.
215. Spier, R. (2002). The history of the peer-review process. *Trends in Biotechnology*, Volume 20, Issue 8, pp. 357–358
216. Srba, I. and Bielikova, M. (2015). Dynamic group formation as an approach to collaborative learning support. *IEEE transactions on learning technologies*, vol. 8(2), pp. 173-186.
217. Stewart, S. and McLoughlin, C. (2007). Design features of an e-mentoring system for the health professions: Choosing to learn in partnership. *The Australasian Society for Computers in Learning in Tertiary Education (ascilite). ICT: Providing Choices for Learners and Learning*. ISBN 978-981-05- 9579-1 2-5. December 2007, Singapore.; pp. 951-961.
218. Strnad, D. and Guid, N. (2009). A fuzzy-genetic decision support system for project team formation. *Journal of applied soft computing*.

219. Swanson, D.; Case, S. and Van der Vleuten, C. P. M. (1991). Strategies for students' assessment. In Boud D., Feletti G., eds. *The challenge of problem-based learning*. London: Kogan Page, pp. 260-274.
220. SWoRD Research (2014). SWoRD Research. URL: <https://sites.google.com/site/swordlrdc/> [Accessed:26-July-2014]
221. Tanimoto, S. L. (2007). The Squeaky Wheel Algorithm: Automatic grouping of students for collaborative projects. Workshop on personalisation in learning environments at individual and group level in conjunction with 11th Int'l Conf. User Modeling, pp. 79-80.
222. Taylor, E. W. (2008). Transformative learning theory. *New directions for adult and continuing education*, number 119, pp. 5-15, Wiley Periodicals Inc.
223. Terenzini, P. T. and Pascarella, E. T. (1994). Living with Myths: Undergraduate Education in America. *Change*, volume 26(1), pp. 28-33.
224. Tobar, C. M. and de Freitas, R. L. (2007). A support tool for student group definition. *The 37th ASEE/IEEE Frontiers in Education Conference*, October 10-13, Milwaukee, WI, USA.
225. Tonkin, E. (2005). Making the case for a wiki. URL: <http://www.ariadne.ac.uk/issue42/tonkin> [Retrieved: 16-April-2014].
226. Tournament organization. URL: <http://sss.wrdsb.ca/sites/sss.wrdsb.ca/files/Tournament%20Organization.pdf> [Accessed: 31-July-2014].
227. Turner, S.; Perez-Quinones, M.A., Chase, J.: Peer review in CS2: conceptual learning. In: SIGCSE 2010 Proceedings of the 41st ACM technical symposium on computer science education, pp. 331–335
228. Van Rooyen, S.; Godlee, F.; Evans, S.; Black, N. and Smith, R. (1999). Effect of open peer review on quality of reviews and on reviewers' recommendations: a randomised trial. *British Medical Journal*, Volume 318, pp. 23-27.

229. Van Rooyen, S.; Godlee, F.; Evans, S.; Smith, R. and Black, N. (1998). Effect of blinding and unmasking on the quality of peer review: a randomized trial. *JAMA*, Volume 280(3): pp.234-237.
230. Verhagen, P. (2006). *Connectivism: A new learning theory?* Retrieved from <http://elearning.surf.nl/e-learning/english/3793>
231. Vivacqua, A. and Lieberman, H. (2000). Agents to assist in finding help. Proceedings of the ACM Conference on Computers and Human Interface (CHI-2000), the Hague, Netherlands, 65–72.
232. Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. M. Cole, V. John-Steiner, S. Scribner and E. Souberman (Eds.). Cambridge, Massachusetts: Harvard University Press.
233. Vygotsky, L.S.: *Mind in Society: The Development of Higher Psychological Processes*. Harvard University Press, Cambridge (1978)
234. Wagner, W. And Steinzor, R. (2006). *Rescuing Science from Politics Regulation and the Distortion of Scientific Research*. Cambridge University Press, 1st edition
235. Wang Y. and Vassileva J. (2007) Toward Trust and Reputation Based Web Service Selection: A Survey. In *International Transactions on Systems Science and Applications*, vol. 3 (2), pp. 118-132.
236. Wang, D. Y., Lin, S. S. J., and Sun, C. T. (2007). DIANA: A computer-supported heterogeneous grouping system for teachers to conduct successful small learning groups. *Computers in human Behaviors*, 23 (4), 1997-2010.
237. Wang, F. and Burton, J. (2010). Collaborative learning problems and identity salience: a mixed methods study. *Journal of educational technology development and exchange*, vol. 3(1), pp. 1-12.
238. Waycott, J.; Sheard, J.; Thompson, C. And Clerehan, R. (2013) Making students' work visible on the social web: A blessing or a curse? *Journal of computers and education*, volume 68(2013), pp. 86-95

239. Weber, E. J.; Katz, P. P.; Waeckerle, J. F. And Callaham, M. L. (2002). Author perception of peer review: Impact of review quality and acceptance on satisfaction. The journal of American medical association (JAMA), volume 28, number 21, pp. 2790-2793.
240. Weber, E. J.; Katz, P. P.; Waeckerle, J. F. And Callaham, M. L. (2002). Author perception of peer review: Impact of review quality and acceptance on satisfaction. The journal of American medical association (JAMA), volume 28, number 21, pp. 2790-2793.
241. Webster, A. and Vassileva, J. (2006). Linking in Lurkers: The Comtella Discussion Forum. In Workshop on the Social Navigation and Community based Adaptation Technologies (2006)
242. Weisstein, Eric W. (2014). "Bipartite Graph." From *MathWorld*--A Wolfram Web Resource. <http://mathworld.wolfram.com/BipartiteGraph.html> [Accessed: 28-July-2014].
243. Wessner, M. and Pfister, H. R. (2001). Group formation in computer-supported collaborative learning. International ACM SIGGROUP conference on supporting group work (GROUP '01), pp. 24-31.
244. Whyte, T. (2011). Connectivism. URL: <http://etec.ctlt.ubc.ca/510wiki/Connectivism> [Accessed: 18-December-2014].
245. Wichmann, A. and Rummel, N. (2013). Improving revision in wiki-based writing: coordination pays off. Journal of Computers and Education, volume 62(2013), pp. 262-270.
246. Wikipedia (2014). Wiki. URL: <http://en.wikipedia.org/wiki/Wiki> [Accessed: 27-June-2014]
247. Wikipedia (2014). Wikipedia: Editcountitis URL: <http://en.wikipedia.org/wiki/Wikipedia:Editcountitis> [Retrieved: 17-April-2014]
248. Wiles, J. A. and Cornwell, B. (1990). A Review of Methods Utilized in Measuring Affect, Emotion in Advertising. Current Issues and Research in Advertising, Volume 13(2), pp. 241-275.

249. Winter, M. and McCalla, G. (2003). An Analysis of Group Performance in Terms of the Functional Knowledge and Teamwork Skills of Group Members. Proc. Int. Conference on Artificial Intelligence in Education (AIED 2003), Sydney, Australia, July 2003, pp. 261-268.
250. Wong, A. T. and Premkumar, K. (2007). An Introduction to Mentoring Principles, Processes and Strategies for Facilitating Mentoring Relationships at a Distance. URL: <http://www.usask.ca/gmcte/drupal/?q=resources> [Accessed: 05-February-2013].
251. Xiong, W. and Litman, D. (2011). Automatically Predicting Peer-Review Helpfulness. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, pages 502–507, Portland, Oregon, June 19-24, 2011.
252. Xiong, W., Litman, D., Schunn, C.: Assessing reviewers' performance based on mining problem localization in peer-review data. In: Proceedings of the Third International Conference on Educational Data Mining (EDM2010), pp. 211–220 (2010)
253. Yamagishi, T. and Kikuchi, M. (1999). Trust, gullibility and social intelligence. *Asian Journal of Social Psychology*, volume 2, pp. 145-161.
254. Yannibelli, V. and Amandi, A. (2012). A deterministic crowding evolutionary algorithm to form learning teams in a collaborative learning context. *Expert systems with applications*, vol. 39, pp. 8584-8592.
255. Yu, H.; Liu, C. and Zhang, F. (2011). Reciprocal Recommendation Algorithm for the Field of Recruitment. *Journal of Information and Computational Science*, Volume 8(16), pp. 4061-4068.
256. Zhang, X., Soh, L. K., Liu, X., & Jiang, H. (2005). Intelligent Collaborating Agents to Support Teaching and Learning. *IEEE International Conference on Electro Information Technology*, May 22-25, Lincoln, NE.

APPENDIX A

Part 1: Total gain from the groups

<i>Algo 1</i>			
<i>Data-points</i>	<i>Uniform</i>	<i>Normal</i>	<i>Pareto</i>
120	19.91118	8.751431	0.089018
240	52.35171	43.901	0.058487
600	67.6717	99.58519	0.446985
840	122.9116	28.9644	0.126499
1080	179.1263	181.9588	0.21839

<i>Algo 2</i>			
<i>Data-points</i>	<i>Uniform</i>	<i>Normal</i>	<i>Pareto</i>
120	14.2831	9.31407	0.147476
240	43.56519	36.03483	0.988404
600	66.19291	51.0871	1.213843
840	95.22513	75.62949	0.221821
1080	120.9173	96.78137	1.259822

<i>Algo 3</i>			
<i>Data-points</i>	<i>Uniform</i>	<i>Normal</i>	<i>Pareto</i>
120	29.1852	28.04075	1.389488
240	89.1671	88.62607	58.08821
600	148.9047	147.3617	62.12806
840	208.9452	208.6212	15.07549
1080	269.4014	267.875	78.1347

<i>Algo 4</i>			
<i>Data-points</i>	<i>Uniform</i>	<i>Normal</i>	<i>Pareto</i>
120	29.1852	28.50576	1.389488
240	89.54695	89.19421	58.08821
600	149.1774	147.3617	62.12806
840	209.2816	208.6212	17.38614
1080	269.4014	268.8121	90.3795

Part 2: Total time consumption of each algorithm

<i>Algo 1</i>			
<i>Data-points</i>	<i>Uniform</i>	<i>Normal</i>	<i>Pareto</i>
120	0.169441	0.01564	0.00065
240	0.001344	0.001293	0.001225
600	0.001862	0.00185	0.002324
840	0.002993	0.002866	0.004039
1080	0.003159	0.003244	0.003412

<i>Algo 2</i>			
<i>Data-points</i>	<i>Uniform</i>	<i>Normal</i>	<i>Pareto</i>
120	0.094102	0.001009	0.001119
240	0.003288	0.002284	0.002876
600	0.0033	0.004386	0.005246
840	0.009194	0.006027	0.007105
1080	0.01149	0.007308	0.007629

<i>Algo 3</i>			
<i>Data-points</i>	<i>Uniform</i>	<i>Normal</i>	<i>Pareto</i>
120	0.470364	0.006053	0.007168
240	0.020999	0.018417	0.021211
600	0.027618	0.019333	0.04022
840	0.047417	0.035936	0.037958
1080	0.078794	0.031633	0.047205

<i>Algo 4</i>			
<i>Data-points</i>	<i>Uniform</i>	<i>Normal</i>	<i>Pareto</i>
120	0.005385	0.000614	0.000601
240	0.001235	0.001215	0.00115
600	0.00174	0.001667	0.002312
840	0.002639	0.002664	0.00254
1080	0.003122	0.003164	0.003203

Part 3: Total space consumption of each algorithm

<i>Algo 1</i>			
<i>Data-points</i>	<i>Uniform</i>	<i>Normal</i>	<i>Pareto</i>
120	3.53E+08	3.54E+08	3.54E+08
240	3.54E+08	3.54E+08	3.54E+08
600	3.54E+08	3.54E+08	3.54E+08
840	3.54E+08	3.55E+08	3.55E+08
1080	3.55E+08	3.55E+08	3.55E+08

<i>Algo 2</i>			
<i>Data-points</i>	<i>Uniform</i>	<i>Normal</i>	<i>Pareto</i>
120	3.54E+08	3.54E+08	3.54E+08
240	3.54E+08	3.54E+08	3.54E+08
600	3.54E+08	3.54E+08	3.54E+08
840	3.54E+08	3.55E+08	3.55E+08
1080	3.55E+08	3.55E+08	3.55E+08

<i>Algo 3</i>			
<i>Data-points</i>	<i>Uniform</i>	<i>Normal</i>	<i>Pareto</i>
120	3.54E+08	3.54E+08	3.54E+08
240	3.54E+08	3.54E+08	3.54E+08
600	3.54E+08	3.54E+08	3.54E+08
840	3.54E+08	3.55E+08	3.55E+08
1080	3.55E+08	3.55E+08	3.55E+08

<i>Algo 4</i>			
<i>Data-points</i>	<i>Uniform</i>	<i>Normal</i>	<i>Pareto</i>
120	3.53E+08	3.54E+08	3.54E+08
240	3.54E+08	3.54E+08	3.54E+08
600	3.54E+08	3.54E+08	3.54E+08
840	3.54E+08	3.55E+08	3.55E+08
1080	3.55E+08	3.55E+08	3.55E+08

APPENDIX B

Exit Questionnaire (CMPT 408 – Winter 2013 & 2014 / CMPT 880 – Fall 2013/Winter 2014)

This is the final survey to evaluate your overall experience with the peer-review coursework used in the CMPT 408 class. The survey contains 29 questions and requires about 20 minutes to complete. Thank you for your time and feedback.

* Required

1. Please enter your name *

2. As an author, what is your general impression of the reviewers? *

(Please select all relevant options)

- Competent
- Constructive
- Provided detailed suggestions
- Provided useful correction to style and grammar
- Too polite
- Not too helpful
- Not too thorough or detailed
- Too negative
- Lacked substance

Didn't always understand the points I was trying to make

3. Did you notice any difference in the reviews you got over time? *

- Improved
- Got worse
- Hard to say

4. Do you think the competence of the reviewers improved over time? *

- Yes
- No
- Hard to say

4b. Comment



5. Do you think your essays were reviewed by the same set of reviewers over the weeks? *

- Yes
- No
- Hard to say

5b. Comment



6. How do you think the assignment of essays to reviewers was done? *

- Based on some principle
- Randomly
- Hard to say

6b. Comment



7. If you chose "Based on some principle", can you guess the principle that the assignment was based on?



8. Please rate the reviews you got from the following essays on the scale of 1 to 10 *

(1- very poor, 10 – excellent)

1 2 3 4 5 6 7 8 9 10

Privacy in Social Networks	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Intellectual Property	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Games and Gamification	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The Future of Education	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Cybercrime and Cyber Security	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Evaluating and Controlling Technology	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Final Paper	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

9. Why did you not evaluate the review you got from the 4th essay? *

(4th essay was "The Future of Education" from 11th to 17th February 2013)



10. Please rank the following, on the scale of 1 to 10, according to their degree of relevance in the choice of the rating you gave to the review of your essay *

(1- very poor, 10 – excellent)

	1	2	3	4	5	6	7	8	9	10
Friendliness	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Good points raised	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Useful corrections to style and grammar	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Suggestions on good ways of expressing ideas	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Lack friendliness	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Lack good points in the review	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Lack useful suggestion on ways of expressing ideas	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

11. As a reviewer, what factor(s) was more important for you to improve your reviews? *

- Feedback from the authors of the essays you reviewed
- Marks given by the marker
- Just trying to be helpful to my colleague

Neither

12. If neither of the above was a motivation for you, what motivated you to give good reviews?

13. What do you think of the quality of the reviews you wrote over the weeks? *

Improved

Stayed the same

Got worse

13b. Comment

14. Can you compare the feedback from the authors of the essays you reviewed and the marks given by the marker? *

Similar

Not similar

- Hard to say: some were similar, some were not

15. In the case that you answered "Not similar", or "Hard to say:", which feedback did you think was fairer?

- Feedback from authors
- Marks given by the marker

15b. Comment



16. Did you check the reviews provided by the other reviewers of the same essays that you reviewed? *

- Yes
- No

16b. Comment



17. If "Yes", did you feel that you learned something from the reviews given by others?

- Yes

No

17b. Comment



18. Did you ever engage in a discussion with the other reviewers about the essays you reviewed? *

Yes

No

19. If "Yes", what did you discuss with other reviewers?

(Please select all relevant options)

Understanding the essay

Emphasizing specific points

Grammar

Style of writing

Other:

20. If "No", why didn't you discuss?

21. How did you decide what rating to give to an essay based on the issues you raised in the review? *

22. Do you think your writing skill has improved as a result of this type of coursework? *

Yes

No

Other:

22b. Comment

23. Do you think your skills as a critic/reviewer have improved as a result of this type of coursework? *

Yes

No

Other:

23b. Comment



24. Do you think this kind of coursework (peer-review) should continue in CMPT 408 *

Yes

No

25. Do you think this kind of coursework (peer-review) should be used in other classes? *

Yes

No

26. If "Yes", please suggest other classes



27. If "No", please give reason(s)

A vertical scrollbar with a track, a thumb, and arrowheads at the top and bottom.

28. What do you think of SWORD as a tool to support the peer review process?

A vertical scrollbar with a track, a thumb, and arrowheads at the top and bottom.

29. Please, provide any additional comments or suggestions, if you have some, below:

A vertical scrollbar with a track, a thumb, and arrowheads at the top and bottom.

Submit

APPENDIX C

Part 1: Informed Consent Form – MPAcc – Fall 2015 / CMPT 408 – Winter 2016

You are invited to participate in a study entitled “Adapting Online Collaborative Learning Tools (Peer-Review Systems) to Support Group Peer Mentorship”.

Please read this form carefully, and feel free to ask questions you might have.

Researchers: Julita Vassileva, Department of Computer Science, jiv@cs.usask.ca (966-2073)
Fred Phillips, Edwards Schools of Business, phillips@edwards.usask.ca (966-8401)
Oluwabunmi Adewoyin, Department of Computer Science, ofa531@mail.usask.ca
(966-8654)

The purpose of the study is to show how online collaborative tools like peer review conference management systems can be improved to support group peer mentorship systems. Findings from this study will be used to improve online collaboration for e-learning platforms, specifically group peer mentorship systems.

In this study, you will respond to two accounting cases and provide peer feedback on three of your peers` case responses, using a peer review conference management system. In addition, you will be asked to complete a personality survey and a post-study survey that asks about your experience with the review system. The case analyses and reviews will require approximately 90 minutes to complete, and the final review-experience survey will require 15 minutes to complete. The personality survey will require about 5 minutes and can be completed now. This survey is hosted by fluid Survey, a USA owned company. See the following for more information on Fluid Survey Data Privacy in Canada. The research data will be stored on a password-protected computer system and will be available only to the investigators. Personally identifying information will be destroyed upon completion of data collection, and pseudonyms will be used to refer to the participants. All data will be kept for a minimum of five years upon the completion of this study. Aggregate results will be used in a thesis and articles published in

conferences and journals. Any information that can be linked to a specific participant will be removed or altered.

Your participation is voluntary, and you may withdraw from the study for any reason, at any time, before the data has been pooled together and anonymized, which will happen after the end of the data collection process. After that point it won't be possible to identify your data, so unfortunately, it won't be possible to withdraw from the study. Before that point, you can withdraw without penalty of any sort, by sending an email to the student investigator on ofa531@mail.usask.ca. If you choose to withdraw, you can still continue to use the system, but no data will be stored about your review ratings and you will not be asked to fill any questionnaires and the data you have contributed will be destroyed at your request. When answering the questionnaires, you may refuse to answer individual questions. If you have any questions concerning the study, please feel free to contact the researchers at the numbers provided above. This study has been approved on ethical grounds by the University of Saskatchewan Behavioral Research Ethics Board on 13th August 2015. Any questions regarding your rights as a participant may be addressed to that committee through the Ethics Office (966-2084). Out of town participants may call toll free (888) 966-2975. You may find out about the results of the study through MADMUC Lab website (<http://madmuc.usask.ca/>) or by contacting the researchers.

I have read and understood the description provided above; I have been provided with an opportunity to ask questions and my questions have been answered satisfactorily. I consent to participate in the study described above, understanding that I may withdraw this consent at any time.

Yes

No

Continue

Part 2: Rotter`s Interpersonal Trust – Distrust Scale

This survey measures your interpersonal trust / distrust tendency, as a preliminary test in this study. Please indicate the degree to which you agree or disagree with each statement by the following scale:

1=strongly agree

2=mildly agree

3 = agree and disagree equally

4 = mildly disagree

5 = strongly disagree

Once you have completed the 25 items, click SUBMIT

Please indicate the pseudonym you wish to be known for the peer review study

Please enter the email address that you will like to be contacted for this study

Strongly agree	Mildly agree	Agree and disagree equally	Mildly disagree	Strongly disagree
----------------	--------------	----------------------------	-----------------	-------------------

1. Hypocrisy is on the increase in our society.

2. One is better off being cautious when dealing with strangers until they have provided evidence that they are trustworthy.

Strongly agree	Mildly agree	Agree disagree equally	and	Mildly disagree	Strongly disagree
-------------------	--------------	------------------------------	-----	--------------------	----------------------

3. This country has a dark future unless we can attract better people into politics.

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
-----------------------	-----------------------	-----------------------	-----------------------	-----------------------

4. Fear and social disgrace or punishment rather than conscience prevents most people from breaking the law.

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
-----------------------	-----------------------	-----------------------	-----------------------	-----------------------

5. An honor system in which teachers would not be present during exams would probably result in increased cheating.

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
-----------------------	-----------------------	-----------------------	-----------------------	-----------------------

6. Parents usually can be relied on to keep their promises.

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
-----------------------	-----------------------	-----------------------	-----------------------	-----------------------

7. The United Nations will never be an effective force in keeping world peace.

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
-----------------------	-----------------------	-----------------------	-----------------------	-----------------------

Strongly agree	Mildly agree	Agree disagree equally	and	Mildly disagree	Strongly disagree
-------------------	--------------	------------------------------	-----	--------------------	----------------------

8. The judiciary is a place where we can all get unbiased treatment.

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
-----------------------	-----------------------	-----------------------	-----------------------	-----------------------

9. Most people would be horrified if they knew how much of the news that the public hears and sees is distorted.

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
-----------------------	-----------------------	-----------------------	-----------------------	-----------------------

10. It is safe to believe that in spite of what people say most people are primarily interested in their own welfare.

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
-----------------------	-----------------------	-----------------------	-----------------------	-----------------------

11. Even though we have reports in newspapers, radio, TV, and the Internet, it is hard to get objective accounts of public events.

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
-----------------------	-----------------------	-----------------------	-----------------------	-----------------------

12. The future seems very

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
-----------------------	-----------------------	-----------------------	-----------------------	-----------------------

Strongly agree	Mildly agree	Agree disagree equally	and	Mildly disagree	Strongly disagree
-------------------	--------------	------------------------------	-----	--------------------	----------------------

promising.

13. If we really knew what was going on in international politics, the public would have reason to be more frightened than they now seem to be.

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
-----------------------	-----------------------	-----------------------	-----------------------	-----------------------

14. Most elected officials are really sincere in their campaign promises.

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
-----------------------	-----------------------	-----------------------	-----------------------	-----------------------

15. Many major national sports contests are fixed in one way or another.

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
-----------------------	-----------------------	-----------------------	-----------------------	-----------------------

16. Most experts can be relied upon to tell the truth about the limits of their knowledge.

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
-----------------------	-----------------------	-----------------------	-----------------------	-----------------------

17. Most parents

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
-----------------------	-----------------------	-----------------------	-----------------------	-----------------------

Strongly agree	Mildly agree	Agree disagree equally	and	Mildly disagree	Strongly disagree
-------------------	--------------	------------------------------	-----	--------------------	----------------------

can be relied upon to carry out their threats of punishments.

18. Most people can be counted on to do what they say they will do.

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
-----------------------	-----------------------	-----------------------	-----------------------	-----------------------

19. In these competitive times one has to be alert or someone is likely to take advantage of you.

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
-----------------------	-----------------------	-----------------------	-----------------------	-----------------------

20. Most idealists are sincere and usually practice what they preach.

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
-----------------------	-----------------------	-----------------------	-----------------------	-----------------------

21. Most salesmen are honest in describing their products.

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
-----------------------	-----------------------	-----------------------	-----------------------	-----------------------

22. Most students in school would not cheat even if they were sure they could get away

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
-----------------------	-----------------------	-----------------------	-----------------------	-----------------------

Strongly
agree

Mildly agree

Agree
disagree
equally

and

Mildly
disagree

Strongly
disagree

with it.

23. Most repairmen
will not
overcharge, even if
they think you are
ignorant of their
specialty.

24. A large share of
accident claims
filed against
insurance
companies are
phony.

25. Most people
answer public
opinion polls
honestly.

Submit

Part 3: Back-evaluation Survey (MPAcc – Fall 2015 / CMPT 408 – Winter 2016)

Please enter your pseudonym.

Please provide the pseudonyms of the three reviewers that reviewed your work. Kindly follow this nomenclature in questions 1, 3, 5, 7, 9 and 11.

e.g. Review1=Pseudonym

Review2=Pseudonym

Review 3 = Pseudonym

1. As an author, please indicate the HELPFULNESS of each review you received on your work.

Please rate from 1 to 5 (1 being very unhelpful and 5 being very helpful)

	1	2	3	4	5
Review 1	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Review 2	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Review 3	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

2. Please comment on what influenced the rating that you gave each review in question 1.

3. Please indicate how DETAILED you think the review of your work was.

Please rate from 1 to 5 (1 being not detailed and 5 being very detailed)

	1	2	3	4	5
Review 1	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Review 2	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Review 3	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

4. Please comment on what influenced the rating that you gave each review in question 3.

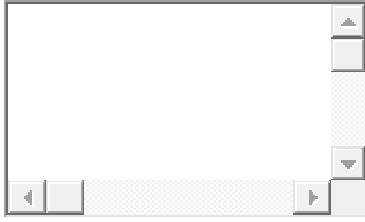


5. How would you rate the CONSTRUCTIVENESS of the reviews of your work?

Please rate from 1 to 5 (1 being not constructive and 5 being highly constructive)

	1	2	3	4	5
Review 1	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Review 2	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Review 3	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

6. Please comment on what influenced the rating that you gave each review in question 5.



7. Kindly rate the POLITENESS of the reviews that you received.

Please rate from 1 to 5 (1 being not polite and 5 being very polite)

	1	2	3	4	5
Review 1	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Review 2	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Review 3	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

8. Please comment on what influenced the rating that you gave each review in question 7.



9. Do you think the reviewers assigned to your work understood the points you were trying to raise in your submitted work?

	Yes	No
Review 1	<input type="radio"/>	<input type="radio"/>
Review 2	<input type="radio"/>	<input type="radio"/>
Review 3	<input type="radio"/>	<input type="radio"/>

10. Please comment on your answer for each review in question 9.

11. Please provide an overall rating of each review that you received.

	Very poor	Poor	Satisfactory	Good	Very good
Review 1	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Review 2	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Review 3	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

12. Please comment on what influenced the overall rating that you gave each review in question 11.

Part 4: Peer Review Session Survey (MPAcc – Fall 2015 / CMPT 408 – Winter 2016)

Question 1: Please enter your pseudonym.

Question 2: As a result of the reviews you received, will you be able to make significant improvement in your work?

Please select the option that best describes each of the reviews that you received

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
Review 1	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Review 2	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Review 3	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Please provide the pseudonyms of the reviewers in question 2.

e.g Review 1 = Pseudonym
 Review 2 = Pseudonym
 Review 3 = Pseudonym

▲

■

▼

◀

▶

Question 3: Do you think the peer review assignment helped to improve your critique skill as a reviewer?

Yes

No

Question 4: If Yes, by how much do you think the peer review assignment has helped your critique skill?

Please rate how much you think the peer review assignment has helped your critique skill by selecting the appropriate star below.

1 2 3 4 5


Question 5: How would you rate your learning from being the reviewer of your peers?

Please select the option that best describes your learning experience for each paper you reviewed

	Very negative	Negative	Neutral	Positive	Very positive
Paper 1	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Paper 2	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Paper 3	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Please provide the pseudonyms of the authors of the papers that you reviewed.

e.g Paper 1 = Pseudonym
Paper 2 = Pseudonym
Paper 3 = Pseudonym



Question 6: Did you look at any reviews prepared by the other reviewers who also were assigned the papers that you were assigned?

Yes

No

Please provide the pseudonyms of the reviewers that reviewed the same paper with you.

Please note the the reviewers and their corresponding pseudonyms. Please structure your response as follows:

Paper 1 = pseudonyms 1a, 1b and 1c
Paper 2 = pseudonyms 2a, 2b and 2c
Paper 3 = pseudonyms 3a, 3b and 3c



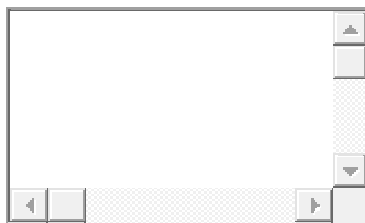
Question 7: If your answer is `Yes` in 6, please consider explaining how it was useful.



Question 8: Please rate your overall experience with this peer review session.

- Very poor
- Poor
- Acceptable
- Good
- Very good


Comments



Submit

APPENDIX D

Behavioural Ethics Approval Certificates

	UNIVERSITY OF SASKATCHEWAN	Behavioural Research Ethics Board (Beh-REB)	<i>Certificate of Approval</i>
PRINCIPAL INVESTIGATOR Julita Vassileva	DEPARTMENT Computer Science	BEH# 12-35	
INSTITUTION(S) WHERE RESEARCH WILL BE CONDUCTED University of Saskatchewan			
STUDENT RESEARCHER(S) Thomas Largillier, Kewen Wu			
FUNDER(S) NATURAL SCIENCES & ENGINEERING RESEARCH COUNCIL OF CANADA (NSERC)			
TITLE Incentivizing High Quality Virtual Collaboration Through Visual Feedback About Conflict and Reputation			
ORIGINAL REVIEW DATE 16-Feb-2012	APPROVAL ON 28-Feb-2012	APPROVAL OF: Application for Approval of Research Protocol (received 23-Jan-2012)	EXPIRY DATE 27-Feb-2013
Full Board Meeting <input type="checkbox"/> Delegated Review <input checked="" type="checkbox"/>			
CERTIFICATION The University of Saskatchewan Behavioural Research Ethics Board has reviewed the above-named research project. The proposal was found to be acceptable on ethical grounds. The principal investigator has the responsibility for any other administrative or regulatory approvals that may pertain to this research project, and for ensuring that the authorized research is carried out according to the conditions outlined in the original protocol submitted for ethics review. This Certificate of Approval is valid for the above time period provided there is no change in experimental protocol or consent process or documents.			
Any significant changes to your proposed method, or your consent and recruitment procedures should be reported to the Chair for Research Ethics Board consideration in advance of its implementation.			
ONGOING REVIEW REQUIREMENTS In order to receive annual renewal, a status report must be submitted to the REB Chair for Board consideration within one month of the current expiry date each year the study remains open, and upon study completion. Please refer to the following website for further instructions: http://www.usask.ca/research/ethics_review/			



Certificate of Re-Approval

PRINCIPAL INVESTIGATOR	DEPARTMENT	Beh #
Julita Vassileva	Computer Science	12-35

INSTITUTION (S) WHERE RESEARCH WILL BE CARRIED OUT

University of Saskatchewan
Saskatoon SK

STUDENT RESEARCHER(S)

Oluwabunmi Adewoyin, Sara Mansouri, Muhammad Mughees

FUNDER(S)

NATURAL SCIENCES & ENGINEERING RESEARCH
COUNCIL OF CANADA (NSERC)

TITLE

Incentivizing High Quality Virtual Collaboration Through Visual Feedback About Conflict and Reputation

RE-APPROVED ON	EXPIRY DATE
17-Jan-2014	16-Jan-2015

- Full Board Meeting
- Delegated Review

CERTIFICATION

The University of Saskatchewan Behavioural Research Ethics Board has reviewed the above-named research project. The proposal was found to be acceptable on ethical grounds. The principal investigator has the responsibility for any other administrative or regulatory approvals that may pertain to this research project, and for ensuring that the authorized research is carried out according to the conditions outlined in the original protocol submitted for ethics review. This Certificate of Approval is valid for the above time period provided there is no change in experimental protocol or consent process or documents.

Any significant changes to your proposed method, or your consent and recruitment procedures should be reported to the Chair for Research Ethics Board consideration in advance of its implementation.

ONGOING REVIEW REQUIREMENTS

In order to receive annual renewal, a status report must be submitted to the REB Chair for Board consideration within one month of the current expiry date each year the study remains open, and upon study completion. Please refer to the following website for further instructions: http://www.usask.ca/research/ethics_review/