
Evaluation of online team-based game development using SNS tools

Akiko Teranishi

New York University Abu Dhabi,
P.O. Box 129188, Abu Dhabi, UAE
Email: akikosakamoto@gmail.com

Minoru Nakayama*

Tokyo Institute of Technology,
152-8552, Tokyo, Japan
Email: nakayama@ict.e.titech.ac.jp
*Corresponding author

Theodor Wyeld

Flinders University,
G.P.O. Box 2100, 5001,
Adelaide, Australia
Email: theodor.wyeld@flinders.edu.au

Mohamad A. Eid

New York University Abu Dhabi,
P.O. Box 129188, Abu Dhabi, UAE
Email: mohamad.eid@nyu.edu

Abstract: Online game development activity using multi-cultural team-based learning (TBL) was examined in order to identify contributing factors to learning activities by analysing the participants' characteristics, communications, and reflections. In this study, 12 collaborating participants from two countries were divided into three teams. The main collaborative task was to develop a game and improve its usability for users. Participants' characteristics and information literacy were measured as parameters for learning activity as well as developing a product for assessment. A learning reflection questionnaire was conducted to evaluate the participants' efficacy, satisfaction and sense of achievement in learning, and any difficulties they experienced. Participants' social media communications were classified into four categories: proposal, permission, encouragement, and acknowledgement, based on the participant's type of communication activity. These relationships were analysed to summarise participants' activities and to develop supporting procedures.

Keywords: team-based learning; TBL; learning reflection; social media interaction; game development; coding; online collaboration; international communication; communication analysis; text analysis; modelling; usability.

Reference to this paper should be made as follows: Teranishi, A., Nakayama, M., Wyeld, T. and Eid, M.A. (2021) 'Evaluation of online team-based game development using SNS tools', *Int. J. Innovation and Learning*, Vol. 29, No. 2, pp.181–206.

Biographical notes: Akiko Teranishi received her BS from the University of California San Diego, majored in Cognitive Science specialised with Human Computer Interaction in 2005. Also, she received her MS from the Tokyo Institute of Technology, majored in Human System Science in 2007. Her Master thesis was 'The proposal and evaluation of situated text input system in mobile phone'. She has been working in Nissan Motor Company as a Vehicle Packaging and Human Machine Interface Engineer since 2007–present. As a researcher, she worked at the New York University Abu Dhabi during 2016–2017, and her current research interests are educational interface design using human computer interaction.

Minoru Nakayama is graduated from the Tokyo Gakugei University in 1983, completed his MEd program in 1985, and received his Doctorate of Engineering from the Tokyo Institute of Technology in 1990. He is currently a Professor at the Information and Communications Engineering, Tokyo Institute of Technology, Japan. His research concerns human visual perception and educational technology.

Theodor Wyeld is on a number of review committees and has published widely. Attracting more than two million in grants and venture capital, he is an inventor of 'thereitis.com'. The patented 'thereitis' technology is the culmination of ten years of research in usability studies on how to better support the sorting of large collections of images. Between 2004 and 2009 he was a research associate with the Australasian Cooperative Research Centre for Interaction Design. He completed his PhD in Cognitive Psychology at the Swinburne University of Technology, Australia.

Mohamad A. Eid received his PhD in Electrical and Computer Engineering from the University of Ottawa, Canada, in 2010. He is currently an Assistant Professor of Electrical Engineering at the New York University Abu Dhabi (NYUAD). He was previously a teaching and research associate at the University of Ottawa from June 2008 until April 2012. He is a recipient of several best paper award in several international conferences and the prestigious ACM Multimedia 2009 Grand Challenge Most Entertaining Award. He has more than 100 conference and journal publications and 6 patents. His academic interests include affective haptics, neurohaptics, and novel haptic interfaces.

1 Introduction

Various types of learning have been conducted in online environments' with participants able to join freely and learn at any time, from almost anywhere. A typical online platform is the massive open online course (MOOC). It has been serving global university online programs for more than a decade (Seaton et al., 2014). Targeted

courses for certain group members are held locally and globally, such as hands-on practice for coding techniques as well as team-based learning (TBL) which encourages participants and promotes their emotional engagement (Wyeld and Nakayama, 2018). Though many organising techniques have been studied (Michaelsen et al., 2002; Gomez et al., 2010), assessment issues for learning performance and participants' attitude to their studies may differ between courses. This style of learning is an increasingly common technique across most subject areas. However, assessment for learning outcomes, and analysis for participant's communication as learning progresses, is sometime ambiguous, as most cases are conducted in an open-ended fashion. Hence, these aspects were examined, in order to promote these types of learning formats, to evaluate learning performance in a formal class setting. This paper outlines an online collaborative project with specific aims and outcomes.

This paper reports on how its authors organised an experiment involving team collaboration for game development in an online environment. Some quantitative metrics of the participants' learning activities were measured and analysed. In particular, the functions of their online message communications was analysed. The temporal appearance of message events was analysed to understand the discussion processes involved (Teranishi et al., 2017b). Further analyses were also required to generalise communication patterns. And, online communications between teams was compared and analysed in order to consider the influence of cultural factors and personal experiences. In this study, relationships between communication behaviour and participants' learning activities are compared across three teams which consist of members from different cultural regions. The effectiveness of team features for the collaborate work was extracted and defined as contributing factors, such as participant's personality, information literacy (IL) and additional personal characteristics for communication-type activities (Bhuasiri et al., 2012; Chen, 2014). Individual communication ability, including their attitude and characteristics, affects their discussion activity and collaboration as a TBL performance exercise. In order to extract these factors for examination, the progress of collaborative discussions and learning activities was carefully analysed. Participant's emotional aspects and learning activities were measured using questionnaires, in addition to discussion activities using lexical analysis.

This paper presents some causal relationships between the measured metrics of the collaborative learning project, and indicate analysing approaches, in order to understand participants' behavioural activities for a small sample of participants.

The following topics are examined in this paper:

- 1 The participants' characteristics are analysed to clarify the effective factors of the participants' discussions, deliverables, and reflections.
- 2 Categorised social media communications among the teams was investigated as a contributing factors for deliverables and the participants' reflections.
- 3 Deliverables between teams are compared in order to evaluate this relationship with the participants' reflections.

2 Related works

Educational methodologies for analysing the learning of computer coding is discussed in Lye and Koh (2014). The skills are recognised as a basic competence for contemporary 21st-century society (Kanbul and Uzunboylu, 2017). Over time, both teaching and learning methodologies for teaching coding have been developed and improved regarding learner's capabilities and changing curriculums (Robins et al., 2003; Lye and Koh, 2014). There has been a paradigm shift of teaching style from one-way teaching to self-directed, collaborative or active learning. Likewise, coding education has similarly introduced various schemes and facilities in accordance with current teaching practices (Lye and Koh, 2014). In particular, recent teaching and learning practices often employ games or competitive settings with peer-to-peer collaborative learning (Combéis et al., 2016; Echeverría et al., 2017).

As a technique to promote intrinsic motivation for learning coding skills, game-based development learning has been introduced to teaching courses (Wu and Wang, 2012; Vahldick et al., 2014). Students are encouraged to engage in game developing. Collaborative learning enhances both the quality of individual learning and game development performance. Team-based peer collaboration helps them achieve their learning performance goals and other skills such as cooperation. The development platform is shared between team members, it provides for a computer supported collaborative learning/work environment (CSCL/W), or TBL environment. The tool discussed in this paper can be shared by participants as an online game development platform (Wyeld and Barbuto, 2014), or computer-supported TBL (CS-TBL) environment. Such environments are used to enhance the participant's learning performance. With peer-to-peer collaboration, face-to-face learning and hands-on learning, the online tool extends their learning activities to include remote partners (Michaelsen et al., 2002; Gomez et al., 2010).

Many web-based learning/teaching systems have been developed and introduced globally (Traphagan et al., 2010). Once established, their sharable online learning/working platforms can be used across institutes or countries, as participants can access them from almost anywhere and at anytime. Typical delivery models for learning materials such as the MOOC (Hill, 2012) are creating a new paradigm in learning method (Open University, 2014). Key learning activities are now increasingly based, not only on individual access to learning material but rather, team discussions including knowledge exchange and collaboration. There are many courses organised by interested parties, and group discussions about online work across various regions which contribute to learning and evaluation cited in Klemmer (2016). Hence, learning activities are beginning to span cross-cultural learning environments. Cross-cultural learning environments can enhance learners' general knowledge and enrich their learning experiences. For example, previous studies have used cross-cultural TBL platforms to teach HCI using 3D collaborative virtual environments (Prasolova-Førland et al., 2007). This includes multicultural teamwork between Australians, Norwegians, and Taiwanese members (Wyeld, 2010). These cross-cultural activities are reported to have enhanced students' knowledge and enriched their learning experiences.

In this paper, we reported a case study of diverse cultural backgrounds: UAE and Japan, and studied details by analysing their social communicatios. Learning project management skills is another area being introduced to university learning for students (Campos and ao Correia de Freitas, 2015). Combined, the activities in a cross-cultural

TBL exercise may contribute also to experience in project administration work. Though the benefits of TBL are well represented in the literature, their quantitative assessment on learning performances or the general effectiveness of collaborative learning is often not included. Such analysis would go some way towards evaluating the effectiveness of these learning strategies and the quality of the artefacts produced. An aim of this project was to address this apparent gap in the literature.

Most people are familiar with the online communication tools used in social media. These same tools are now increasingly implicated in some educational practices for enhancing the quality of student learning (Michaelson et al., 2002). They are also implicated in improved performance using social media (Cheston et al., 2013; Rasiah, 2014). Their effectiveness is dependent on implementation design (Manca and Ranieri, 2017). Good results have been reported in the literature for using social media communication platforms. Especially those where participants are engaged in online collaborative tasks that supports the discussion of dialogues on learning to code in a CSCL environment (Erkens and Janssen, 2008). Most significant impacts are reported as reflections on participants' motivations (Rasiah, 2014).

Learning activity assessment is often focused on participants' communication activities and emotional factors such as satisfaction, in addition to the conventional learning performance factors. As impactful measurable factors, participants' reflections, satisfaction and achievement in learning are analysed here to evaluate their contribution to educational effectiveness in a CSCL environments. Approaches include: emotional factors (Jung et al., 2002; Paechter et al., 2010; Moore and Kearsley, 2011), online communication content (Marra et al., 2004), and factors in participants' attitude and their characteristics (Chen and Caropreso, 2004; Ogot and Okudan, 2006; Nakayama et al., 2008).

Discourse analysis is included in order to extract 'strength of discussion' on the effectiveness of collaborative learning (Armstrong and Hyslop-Margison, 2006). The significant features of participants' dialogues have been analysed (Menekse and Chi, 2019). From this, 'discussion patterns' were extracted as significant categories of communication to determine their contribution to the participants' overall learning performances (Chi and Menekse, 2015).

As a theoretical approach, the contributions of online participatory activities of learners was examined (Hrastinski, 2009). Some possible factors and their relationships are discussed (Shaw, 2013). However, the contributing factors are many, and extraction of their relationships is open to further study (Kebritchi et al., 2016). Possible hypotheses are examined using repeated measures of empirical case studies (Cheng and Chau, 2016). In general, common assessment is focused on participant's satisfaction and relationships with other metrics of learning activities (Bradford and Wyatt, 2010; Shaw, 2013; Cheng and Chau, 2016). Other studies consider the relationship between online activity and learning performance (Prasolova-Førland et al., 2017). This work has tried to create a relationship between learning activity in participant's discussions and their products as learning outcomes using various objective metrics. The repeated measures technique was used to analyse the development processes of participant's learning. The outcomes of this study continue and extend prior studies.

3 Method

In the current study, five Arab and seven Japanese university students participated over a period of three weeks. They were students from the UAE and Japan, and recruited from either information science or computer science classes for the collaboration task. Team sizes were up to five participants, with three teams. In general, it was noted that the Japanese students hesitated to discuss issues online. In contrast, the Arab students were more activities to the online discussions.

There were three teams: Japanese, Arabic and international. Both participants from the UAE and Japan had the necessary abilities in IL and communication skills in English. The same project instructions were provided to both groups at the beginning of the study.

3.1 Collaboration task

3.1.1 Task description for each team

The main topic for the team work task was instructing participants on how to perform an online game development, including information from previous studies as a guide (Prasolova-Førland et al., 2007; Wyeld, 2010). The goal of the game development task was to improve its overall usability by evaluating team member's interactions with each iterative version. A usability scale for evaluation of the product versions was employed. The method of usability measurement for development of software was one of the core aims of the TBL exercise.

3.1.2 Usability development cycles

The duration of the project was three weeks with three development cycles. Each cycle lasted seven days: six days after development and one day after usability testing. In the first cycle, all teams started with a basic game, version 0.1, and developed it further based on the first usability testing results. At the end of each cycle, all participants evaluated another team's game. The evaluation results were collected and sent to each team after the next cycle.

The four implemented versions of games were labelled as follows:

- 1 Version 0.1: Initial game code provided to participant teams.
- 2 Version 1.0: First revisions for week one.
- 3 Version 1.1: Second revisions for week two.
- 4 Version 2.0: Final revisions for week three.

The study flow is illustrated in Figure 2. In this study, formative evaluations using team communications, and summative evaluations using learning reflections, were measured to analyse the interrelationship between factors. The participants' attributes, such as characteristics and IL, were evaluated as a diagnostic evaluation.

3.1.3 System usability scores

The usability of the games as an overall assessment was conducted using system usability scales (SUS) and their scores (Brooke, 1996). The SUS scores were compared between teams as each team's deliverable. Before starting the project, and after each development cycle, the participants individually evaluated the usability of the other teams' games. As there were three development cycles in total, four SUS scores were provided in total. Each team was evaluated by another team using SUS as user feedback. Based on the usability tests, implementation changes were made. The experimenter assigned the participants which (other) team they needed to evaluate; each team was evaluated by two other teams. Before starting the development cycle, the participants evaluated the initial game as well as during the proceeding three development cycles.

3.2 Participants

Twelve students (nine males and three females, between 18 and 24 years age) from two countries participated in this study. They were from UAE and Japan. Participants were divided into three teams depending on their nationalities:

- Domestic 1 team: Arabic team consisted of three Arab students.
- Domestic 2 team: Japanese team consisted of four Japanese students.
- International team: International team consisted of three Japanese students and two Arab students.

All 12 participants signed an informed consent form authorising their social media communication to be used as data for analysis. Participants who completed the team-based work received \$100 compensation.

3.3 Game development environment

The study was conducted using an online 2D mobile phone game editor task. Participants were asked to develop a game as a collaborative exercise with their teammates. Each team started with a basic html and JavaScript browser game, provided by the experimenter. Three different, but equally challenging base games were provided to close from (see Figure 1).

Figure 1 Basic HTML and JavaScript browser games (see online version for colours)

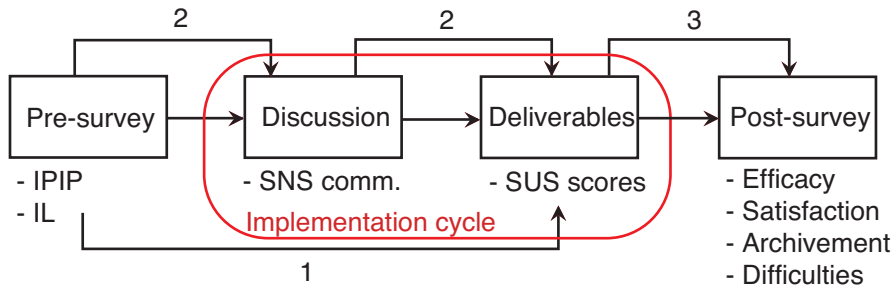


Teams were required to iteratively improve the game with usability testing between guiding improvements. Code changes and updates were based on the usability testing feedback from the evaluating teams. The game was developed using an online game editor provided by the experimenter which supported in browser direct code changes.

The features of each base game are, for the:

- Globe game (domestic team 1): The main character (green circle) scores points by catching and at the same time try to avoid the orange squares which randomly. Once the main character is hit by at least three enemies, the game is over.
- Maze game (domestic team 2): The main (green square) scores points by catching its friends (blue squares) and at the same time tries to avoid the enemies (orange squares) which randomly attack. Once the main character is hit by at least three enemies, the game is over. The movement is restricted by the maze shape.
- Catcher game (international team): The main character (green square) scores points by catching its friends (orange squares) which randomly fall down from the top of the screen, as do enemy shapes (orange squares). Once the main character fails three times (lives shown as red squares on the top right) to catch friends, the game is over.

Figure 2 Conceptual diagram (see online version for colours)



3.4 Characteristics of participants

Before stating the usability development study, the participants were asked to complete the International Personality Item Pool (IPIP) test to measure their personalities (Goldberg, 1999) and an IL test to measure their skills and knowledge (Fujii, 2007). This was done as the characteristics of individual participants may affect their development progress and achievement.

3.4.1 Personality

The personalities of students were measured using the IPIP (2004) inventory. The following five factors are specified from 50 questions: 'extroversion' (IPIP-1), 'agreeableness' (IPIP-2), 'conscientiousness' (IPIP-3), 'neuroticism' (IPIP-4) and 'openness to experience' (IPIP-5).

3.4.2 Information literacy

IL was also measured using 32 question inventories which were defined as eight factors by Fujii (2007). In addition, two meta-factors were extracted as operational skills (IL-1) and attitudes toward IL (IL-2) (Nakayama et al., 2008).

3.5 Monitoring team activity

3.5.1 Discussion via social media

All social media communications and discussions were collected at the end of the development cycle for categorisation and analysis. All teams were required to discuss and reflect on their implementations via online/offline communications. All groups decided to use some form of social media application to interact with teammates. The experimenter did not specify the type of communication they should use. Most of the social media applications had similar functions and features. All three teams decided to use different social media applications. They used variously: Skype messenger, Facebook messenger, and The Line (which is another SNS smart-phone application). Both domestic teams (Japanese and Arabic) were able to communicate without social media, but they decided to use it as one of their communication tools anyway. All teams used a text messaging system, which included the ability to attach files simultaneously, such as screenshots and word/text documents. The Arabic team communicated via Facebook messenger on the first and third cycles. In the second cycle, they met on campus to discuss, therefore no data was collected for this cycle. The Japanese team used Line, which is the most commonly used SNS application in Japan. Line application has a similar functionality with other messengers. The International team decided to use Skype (video) and Skype text messenger. Skype messenger was used as a more precise communication tool to exchange information. Having members from two different countries, with a five-hour time difference meant the teams needed to agree on timing of weekly discussions from the beginning. Arabic teammates had met as classmates before, whereas the Japanese and international teammates had not met their fellow teammates before.

3.5.2 Learning reflection

At the conclusion of the study, individual participants were required to complete a questionnaire reflecting on their learning experiences. The survey consisted of three parts:

- 1 self-efficacy (ten questions) and difficulty of performing the task (ten questions)
- 2 satisfaction of learning
- 3 achievement of learning.

They were self-evaluated at the end of the course to measure learning reflection (Teranishi et al., 2017a, 2017b). In this study, the self-efficacy contained motivational efficacy which was based on previous work (Pintrich and Groot, 1990; Nakayama et al., 2015), and technical efficacy to measure their understanding of the learning activities.

All participants were asked to answer using a five-point mono-polar scale. The question items are shown in Figure 5 for efficacy and Figure 6 for difficulty.

Satisfaction and achievement in learning were also surveyed to measure the participants' level of emotional experience, given as a 100 point score. Their answers included both their expectations at the beginning of the study (defined as initial) and their actual experience (defined as final) based on their learning reflection. In the next section, we examine the differences between the initial and final results, and how they may be affected by their SNS communications.

4 Results

4.1 Survey results

4.1.1 Characteristics and skills

Participant's readiness, such as personality and IL, are compared between the three teams. The results of IPIP and IL scores are summarised in Table 1. These results show that all teams have similar characteristics. In particular, the level of skills and attitudes for IL. This is confirmed by the results of statistical testing using one-way ANOVA which shows there were no significant differences (discussed in the next section).

Table 1 Summary of IPIP and IL tests for each team: mean (SD)

			<i>Domestic1</i>	<i>Domestic2</i>	<i>International</i>
IPIP	1	Extraversion	3.5 (0.3)	3.0 (1.4)	3.2 (1.0)
	2	Agreeableness	4.5 (0.2)	3.5 (0.7)	4.1 (0.4)
	3	Conscientiousness	3.3 (0.3)	2.9 (0.4)	3.0 (0.5)
	4	Emotional stability	3.4 (1.0)	2.6 (1.0)	2.7 (0.2)
	5	Intellect	3.9 (0.8)	3.0 (0.7)	3.5 (0.6)
IL	IL-SF1	Skills	4.0 (0.3)	4.2 (0.3)	4.1 (0.4)
	IL-SF2	Attitude	3.4 (0.5)	3.2 (0.2)	2.8 (0.3)

4.1.2 System usability scales

The averages of the SUS scores are shown in Figure 3. Domestic team 2 achieved a score of 49, whereas the other two teams achieved 66. Additionally, the overall usability scores did not improve significantly for domestic team 1, unlike other teams, where the usability scores increased considerably. The most significant improvement from version 0.1 to 2.0 was performed by domestic team 2.

Relative SUS scores were calculated by subtracting the average of the evaluation scores for other teams from the SUS scores (summarised in Figure 4). The relative SUS scores indicate how much the participants were evaluated relatively compared to their own games.

Both the SUS and the relative SUS scores were not affected by the number of teammates, since domestic team 1, which had four members, performed the best in terms of the SUS scores although domestic team 2 had three members and the international team had five members.

Figure 3 Comparison of SUS's changes across the three teams (see online version for colours)

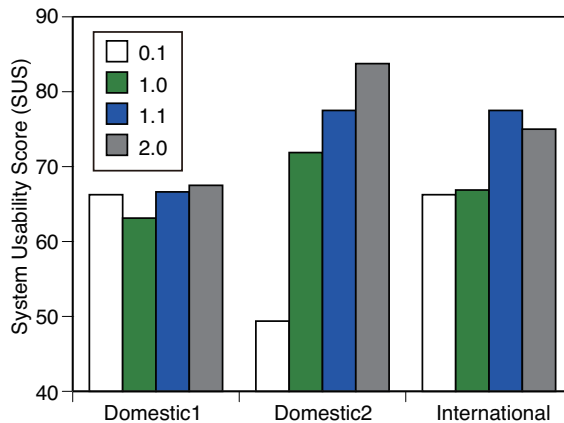
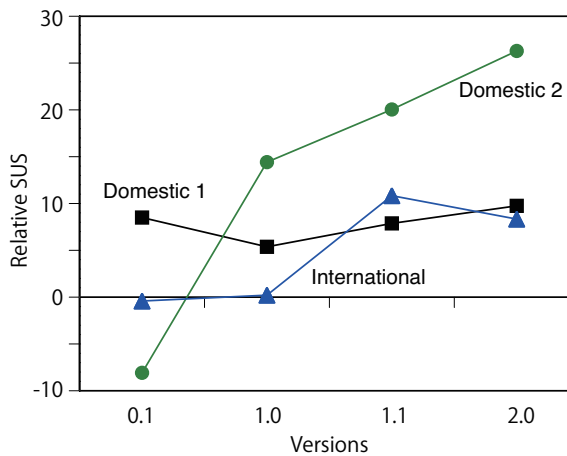


Figure 4 Changes of relative SUS scores along the stages of development (see online version for colours)



4.1.3 Participant's responses

Participants' responses were surveyed using questionnaires as discussed in the section on method. Both questionnaires concerning efficacy and difficulty consisted of ten items. Rather than using factor analysis, cluster analysis using the Ward method (Gatignon, 2014) was employed to summarise the results of these questionnaires into several factors. A dendrogram for efficacy question items is illustrated in Figure 5, where items are summarised as commitment, skills, and vigour. Three components of difficulty, such as communication, project task, and collaboration, are extracted from a dendrogram as a result of the cluster analysis shown in Figure 6.

The results of the mean value of the clustered efficacy are shown in Figure 7. Domestic team 2 had the highest score in both commitment and skills, although domestic team 1 had the highest score in vigour. There were no statistical differences in efficacy between teams.

Figure 5 Clustering for responses of efficacy question items

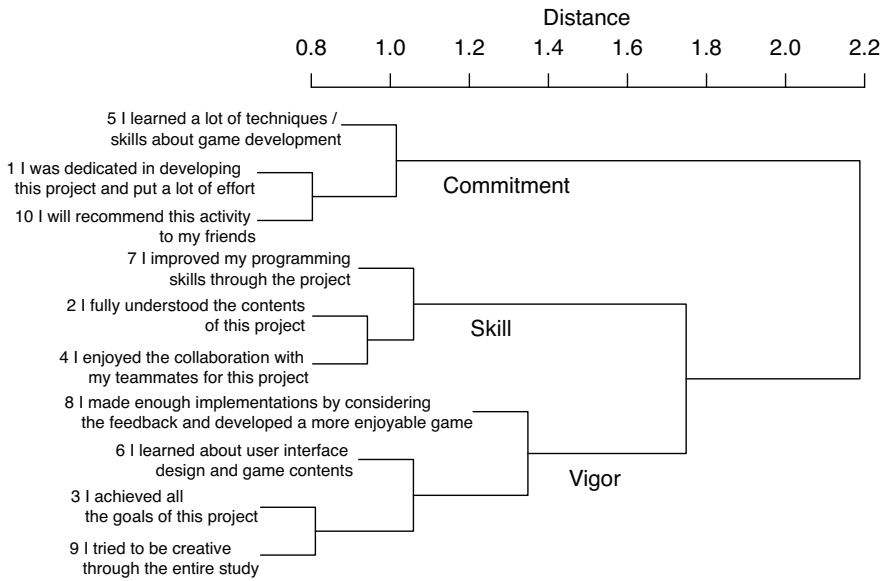
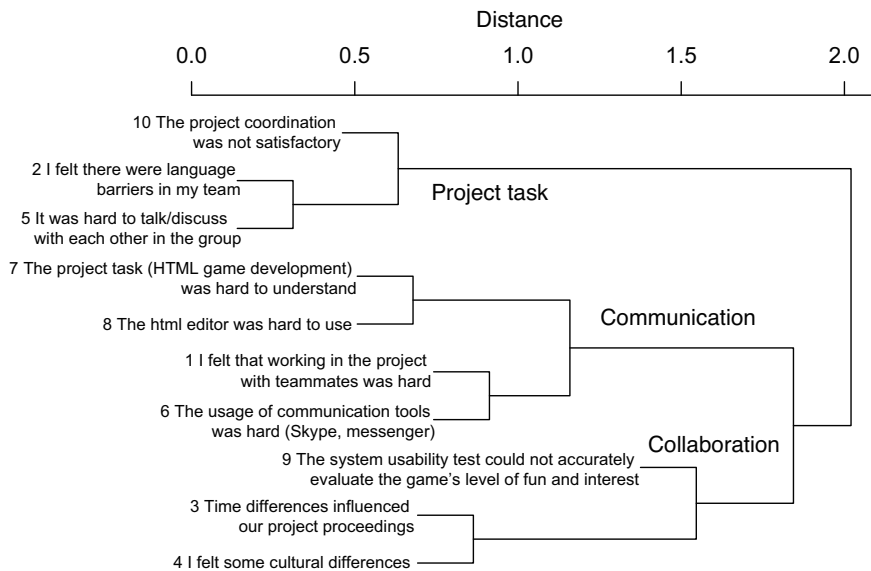


Figure 6 Clustering of responses for difficulty question items



The results of the mean value of the clustered difficulties are shown in Figure 8. The International team experienced the highest difficulty scores among all three aspects. Compared to the other teams, they faced more significant difficulties in terms of the project task (ANOVA; Tukey post hoc, $p < 0.01$). It is considered that these difficulties affected their SUS scores.

Figure 7 Mean scores of efficacy across the three teams (see online version for colours)

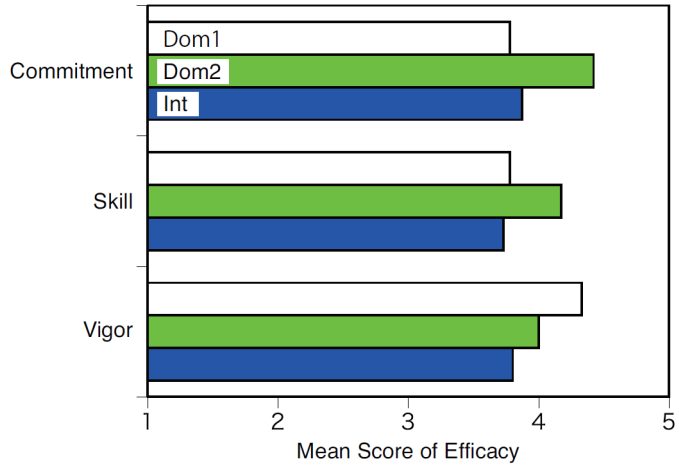


Figure 8 Mean scores of difficulty across the three teams (see online version for colours)

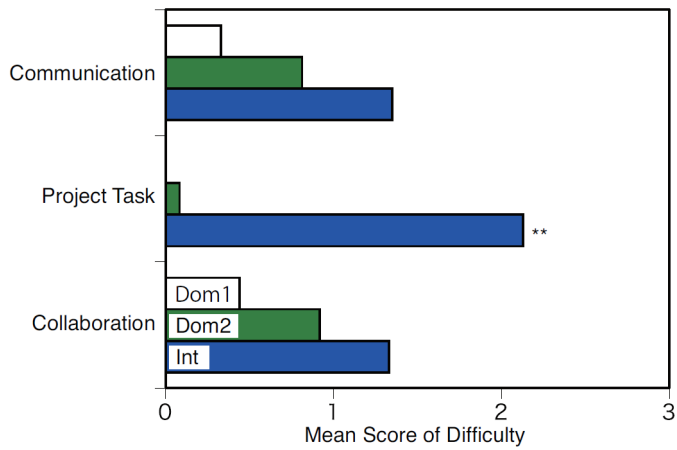


Figure 9 Mean scores of satisfaction across the three teams (see online version for colours)

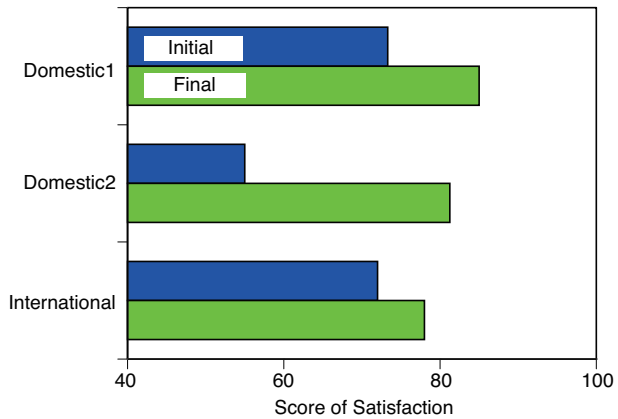
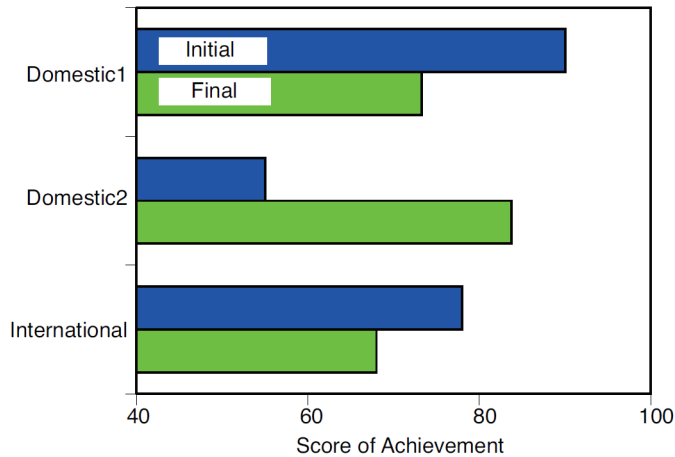


Figure 10 Mean scores of achievement across the three teams (see online version for colours)

Both scores for satisfaction and achievement were compared between the initial and final stages in Figures 9 and 10. The overall tendencies, and degrees of satisfaction, increased during the collaboration work, while degrees of achievement decreased. There were no significant differences between teams for both satisfaction and achievement from the two surveys. The small number of participants may have influenced the test results.

4.2 *Communication analysis*

The total number of social media communications across all teams was 392. The social media communications from all groups were divided into two categories: project related communications (PRC), and non-project related communications (non-PRC).

PRC was categorised into four different types of communication using protocol analysis. Protocol analysis is often used to classify communication and dialogue (Daly et al., 1989; Aleman and Vangelist, 1994). It is often also used to evaluate social media communications (Erkens and Janssen, 2008; Hara et al., 2000). In this study, the social media communications were categorised as: proposal, permission, encouragement, and acknowledgement, with the following criteria:

- Proposal: Dialogues which include information about a new implementation idea.
- Permission: Acceptance against someone's proposal such as 'okay' and 'I think so'.
- Encouragement: Communication where someone encourages other teammates, such as 'we will do our best'.
- Acknowledgement: Notification and acknowledgement where students replied against a teammate's work such as 'thank you' and 'I changed'. It also includes information what she/he did, and thought, based on their discussion.

Protocol analysis was conducted by the experimenters; 88.6% of the classifications were initially matched, and the rest of the unmatched items were decided after

discussion. The numbers for each categorised communication are shown in Table 2. The frequencies of the categorised communications seem to be related to the team activities, therefore the SNS scores are calculated using the weighting coefficients to investigate the relationships with the learning reflections. These were determined by the experimenters based on the importance and the contents of the communications to reflect their qualities. For example, the proposal communication was considered to be the most important communication, and also ten times more significantly important than the acknowledgement communication. As a result, the importance of the communications was resolved in the order of proposal, permission, encouragement, and acknowledgement.

Table 2 Number of categorised social media communications and scores

Communication category	Domestic team 1 (N = 3)			Domestic team 2 (N = 4)			International team (N = 5)		
	Cycle 1	Cycle 2	Cycle 3	Cycle 1	Cycle 2	Cycle 3	Cycle 1	Cycle 2	Cycle 3
	Proposal	17	-	11	4	7	10	1	1
Permission	2	-	1	0	3	0	0	0	0
Encouragement	2	-	0	2	4	5	0	0	1
Acknowledgement	32	-	32	3	16	17	3	5	10
Total N*	53	-	44	9	30	32	4	6	14

Note: *total number of SNS communication.

4.3 Correlation analyses

4.3.1 Relationships between SUS scores and attributes

To address the research questions raised in this paper, the aforementioned results were analysed. Correlational analyses were performed to examine any relationships between the SUS scores and the participants’ attributes, such as characteristics, skills, and attitudes. As a result, we found factors in IPIP were not significantly correlated with the SUS scores. This means participants’ characteristics did not affect their SUS scores. Nevertheless, information skills and attitude factors from IL significantly contributed to the SUS and the relative SUS scores (see Table 3). The result shows that the learners’ attitude was positively correlated with the SUS score ($r = 0.71, p < 0.10$) as well as the relative SUS score ($r = 0.56, p < 0.10$), and that skills were positively correlated with the relative SUS scores in version 0.1 ($r = 0.53, p < 0.10$). However, after version 0.1, attitude was negatively correlated with the relative SUS scores ($r = -0.50$ in version 1.0; $r = -0.61$ in version 1.1, $p < 0.10$). In version 2.0, attitude was also negatively correlated with the SUS scores ($r = -0.56, p < 0.10$) and the relative SUS scores ($r = -0.64, p < 0.10$). These results indicate that the participants who had high IL in terms of skills and attitudes had better usability scores at the beginning of the project, but after version 0.1, high attitude scores caused lower usability scores. This could be interpreted as the feedback from version 0.1 conveyed stricter evaluations across teams. This interpretation appears to be supported by the sample participant survey comment, “the skill of my team was very high, so we could make our game enjoyable.” This indicates that the level of the information skills directly affected the teams performance,

and this appears to be observable at the beginning of the project according to the data analysis also.

4.3.2 Relationships between SUS scores and communications

The relationship between the usability scores and communications among teams was examined in order to better comprehend the effectiveness of team activity (see Table 4). A correlational analysis was performed to examine the relationships between categorised communications and the SUS scores. In cycle 1, the SUS score was correlated to proposal ($r = 0.55, p < 0.10$) and acknowledgement ($r = 0.50, p < 0.10$) as well as the total number of communications ($r = 0.50, p < 0.10$). In cycle 3, the relative SUS score was negatively correlated to the total number of communications ($r = -0.53, p < 0.10$) and proposal communication ($r = -0.51, p < 0.10$). These results indicate that communication, especially proposal and acknowledgement, led to better usability scores at the beginning of the project. To share new ideas and proposals, the brainstorming style discussions seemed to help in the first cycle. However, proposal communication in cycle three seems to have caused an adverse effect on this score. This may be because the third cycle was the last implementation stage, and it might have been too late to discuss new ideas.

Table 3 Correlation coefficients between SUS and IL (N = 12)

Ver.	Usability	Information literacy	
		Skills	Attitude
0.1	SUS	(0.49)	0.71
	r-SUS	0.53	0.56
1.0	SUS	(0.15)	(-0.16)
	r-SUS	(0.08)	-0.50
1.1	SUS	(0.26)	(-0.36)
	r-SUS	(0.21)	-0.61
2.0	SUS	(0.15)	-0.56
	r-SUS	(0.10)	-0.64

Notes: r-SUS: relative SUS; (): not significant coefficient ($p > 0.10$).

Table 4 Correlation coefficients between SUS and communications (N = 12)

Cycle	Usability	Communications				
		Total	Pro	Per	Enc	Ack
1	SUS	0.50	0.55	(0.32)	(0.11)	0.50
	r-SUS	(-0.21)	(-0.21)	(-0.09)	(-0.29)	(-0.19)
2*	SUS	(-0.11)	(-0.16)	(-0.16)	(-0.09)	(-0.06)
	r-SUS	(-0.27)	(-0.32)	(-0.34)	(-0.22)	(-0.21)
3	SUS	(-0.33)	(-0.37)	(-0.22)	(-0.11)	(-0.24)
	r-SUS	-0.53	-0.51	(-0.36)	(-0.09)	(-0.45)

Notes: r-SUS: relative SUS; 2*: N = 9; (): not significant coefficient ($p > 0.10$).

4.3.3 Relationships between SUS scores and reflection

A correlational analysis was conducted to examine the relationships between usability scores and learning reflections. Although there were no significant relationships between SUS scores and learners' efficacy, the relationship between the SUS scores and clustered difficulties (shown in Table 5) identified significant relationships between SUS scores and project task, and SUS scores and collaboration.

Table 5 Correlation coefficients between SUS and scores of difficulty (N = 12)

Ver.	Usability	Score of difficulty		
		Communication	Project task	Collaboration
0.1	SUS	(-0.48)	-0.50	(-0.07)
	rSUS	(-0.30)	(-0.13)	(0.20)
1.0	SUS	(-0.37)	(-0.12)	-0.61
	r-SUS	(0.06)	0.55	(-0.12)
1.1	SUS	(-0.16)	(0.02)	-0.80
	r-SUS	(0.11)	(0.48)	-0.50
2.0	SUS	(0.23)	(0.33)	-0.59
	r-SUS	(0.37)	0.59	(-0.29)

Notes: r-SUS: relative SUS; (): not significant coefficient ($p > 0.10$).

Project task had a negative relationship with the SUS score in version 0.1, but a positive relationship with the relative SUS score in the final version. This may be because project tasks were harder at the beginning, where usability scores were low. However, once the teams started to implement the games' usability, the reduced task difficulty seems to be a contributory factor. This is because the participants struggled with the task as much as they could but achieved their goals in the end. This appears to be supported by a typical comment from the international team: "I had a very enjoyable experience", "it was fun to work with people on the other side of the world." Although the international team felt difficulties the most, in terms of the project task (shown in Figure 8), it shows also that they were able to overcome difficulties in the end.

We examined the negative relationships between the SUS scores and overall collaboration. When the participants felt difficulties arising in collaboration work, the usability scores decreased. In fact, a typical participant comment was: "I enjoyed the project through the teamwork. At the same time, I could improve my programming skill." This tends to indicate that teamwork did indeed enhance participants' skill and enrich their learning experience.

There were no significant relationships between the SUS scores and communication difficulty. This may be because most of the participants were familiar with using social media communication tools and did not feel any difficulty in communicating with their fellow teammates.

4.4 Regression analysis

Communication functions may affect participant's reflections, as was mentioned previously. To confirm the contribution of frequency of communication to participant's reflections scores, multiple regression analysis was conducted to examine the

relationships between reflection variables and the frequencies of the four categories of communication. The results are summarised in Table 6. Table 6 shows the coefficients for each of the three cycles. Significant coefficients are indicated using italics. These relationships produced high R^2 as an effect size.

Table 6 Regression coefficients between participant's reflections and communications (N = 12)

<i>Variables</i>	<i>Cycle</i>	<i>Pro</i>	<i>Per</i>	<i>Enc</i>	<i>Ack</i>	R^2
Ec1	1	(-0.51)	(-0.43)	<i>0.98</i>	(0.21)	0.47
	2	(2.50)	(-2.30)	(0.78)	(-0.58)	0.52
	3	(0.23)	<i>-1.69</i>	<i>0.32</i>	(-0.05)	0.51
Ec2	1	(-0.24)	(-0.30)	(0.36)	(0.13)	0.09
	2	(-1.3)	(-0.67)	(0.00)	(0.92)	0.69
	3	<i>0.27</i>	(-1.04)	(0.27)	(-0.05)	0.54
Ec3	1	(-0.20)	(-0.81)	(0.41)	(0.17)	0.32
	2	<i>-2.50</i>	<i>4.25</i>	<i>-1.58</i>	<i>0.63</i>	0.88
	3	(0.05)	(0.16)	(0.11)	(-0.01)	0.10
Dc1	1	(0.18)	(0.61)	(-0.48)	(-0.20)	0.31
	2	(3.13)	(-3.3)	(1.17)	(-1.13)	0.50
	3	(-0.16)	(-0.07)	(-0.27)	(-0.02)	0.34
Dc2	1	(0.42)	(1.41)	(-0.76)	(-0.40)	0.36
	2	(-0.83)	(-3.3)	(1.22)	(0.67)	0.30
	3	(-0.24)	(-0.13)	(-0.28)	(0.00)	0.26
Dc3	1	(-0.33)	(0.31)	(-0.06)	(0.09)	0.33
	2	(2.00)	(-0.33)	(-0.11)	(-0.83)	0.24
	3	(0.05)	(-0.86)	(-0.05)	(-0.02)	0.10
SAT	1	(-2.44)	(3.73)	(-5.37)	(1.19)	0.07
	2	(-25.0)	(70.0)	<i>-45.0</i>	(10.0)	0.87
	3	<i>9.85</i>	<i>-31.6</i>	(3.16)	(-1.58)	0.79
ACH	1	(-5.6)	(12.7)	(-1.99)	(-0.33)	0.20
	2	(32.5)	(90.0)	(-60.0)	(-16.3)	0.76
	3	<i>14.1</i>	<i>-72.5</i>	(4.33)	<i>-3.73</i>	0.73

Notes: *Italic* indicates significant coefficients, () n.s.

Ec1: commitment, Ec2: skill, Ec3: vigour

Dc1: project task, Dc2: communication, Dc3: collaboration

SAT: change of satisfaction, ACH: change of achievement.

For commitment of the responses of efficacy, the frequency of encouragement posts contributed positively. For skill of the responses of efficacy, the frequency of proposal posts made a positive contribution during the final cycle. For vigour of responses of efficacy, the frequency of posts affected the second cycle.

For the impression of level of difficulty, posted messages did not affect efficacy. The level of difficulty may be an overall impression.

For both satisfaction and achievement in participant's self assessments, the frequency of proposal in the final cycle contributed to their scores. Other frequencies did not positively affect scores.

These results suggest that a specific communication, such as proposal, affected the participants' reflections. Hence, the progress of communication was examined next.

Figure 11 Transition paths and probabilities of the four types of communication

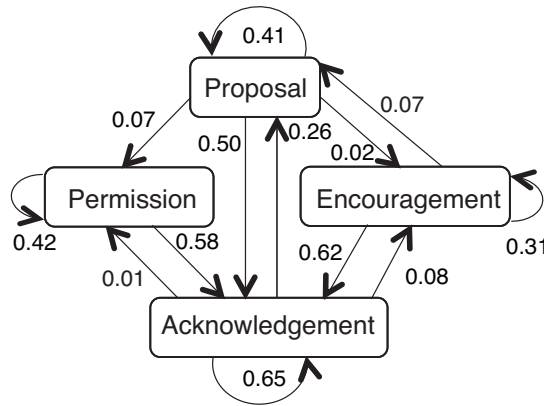
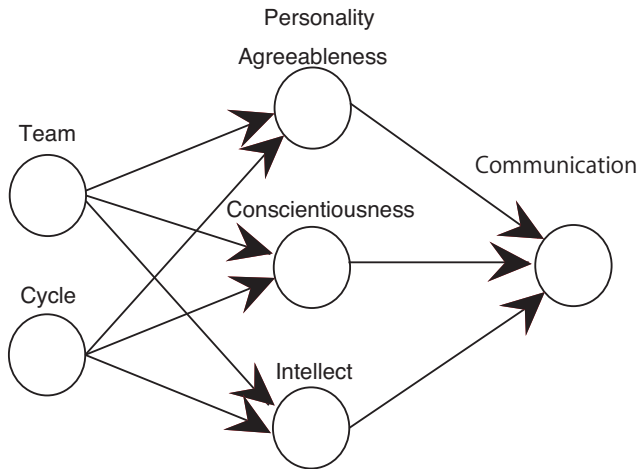


Figure 12 A Bayesian network between communications and personality



4.5 Communication transitions

As the four types of communication appear consecutively, the transitions can be analysed statistically using a Markov model (Teranishi et al., 2017b). A model of all communication data is summarised in Figure 11. Figure 11 shows transitional paths and their probabilities. As the frequencies for permission and encouragement are not large, most communication transitions appear between proposal and acknowledgement. In the previous section, the frequency of proposal communication was identified as a key component of participants’ reflections. The figure shows transitional paths and their probabilities. The production process for proposal is determined using probabilities of transitions to proposal from other post categories. Though the transitional probabilities of most paths to proposal from other categories of posts are low, proposal communications encouraged additional new proposal communications, and thus the transitional probability from proposal to proposal is 0.41. Since proposals led to additional proposals, constructive communication was generated as a result.

These results lead to the next question: “how to generate proposal communication.” The degree to which a participant communicates may depend on the participant’s personality and IL. Factor scores are summarised in Table 1. They show no significant differences between the three teams. The results show that all teams have similar levels of characteristics, such as personality and IL.

Figure 13 A Bayesian network between communications and IL

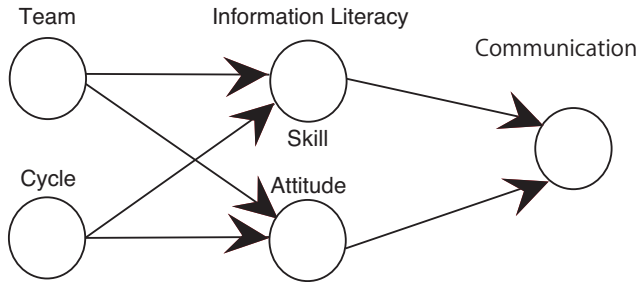


Table 7 Probabilities for communication actions using personality

<i>IPIP-2</i>	<i>IPIP-3</i>	<i>IPIP-5</i>	<i>Pro</i>	<i>Per</i>	<i>Enc</i>	<i>Ack</i>
High	High	High	0.25	0.22	0.22	0.31
Low	High	High	0.30	0.18	0.18	0.35
High	Low	High	0.23	0.23	0.24	0.29
Low	Low	High	0.25	0.23	0.23	0.29
High	High	Low	0.24	0.16	0.15	0.45
Low	High	Low	0.24	0.19	0.24	0.35
High	Low	Low	0.27	0.24	0.24	0.24
Low	Low	Low	0.29	0.18	0.17	0.36

Notes: Pro: proposal, Per: permission, Enc: encouragement, and Ack: acknowledgement; IPIP-2: agreeableness, IPIP-3: conscientiousness, and IPIP-5: intellect.

Discussion communication is often driven by participant’s characteristics. Hence, the impact of their characteristics is examined next. In order to conduct a statistical analysis of characterisation, the Bayesian network technique was employed to measure the contribution of personality and IL. Participants’ characteristics were classified into two levels, such as high or low, with levels classified using their means. Figures 12 and 13 show a Bayesian network model which is influenced by teams and cycles. The model simulates the appearance of communication categories in regards to levels of factor scores for personality or IL. The probability of appearance of posts in the four communication categories can be defined as the probability of when either personality or IL is introduced with IPIP-2, IPIP-3, IPIP-5 for personality, and a probability of an event derived from the IL scores, IL-1 and IL-2. Calculations were made using a R package for Bayesian network (Bottcher and Dethlefsen, 2015).

Results of probabilities for personality are summarised in Table 7. In the table, the left columns show combinations of personality levels. In this case, there are four categories, and the chance of probability approaches 0.25. When agreeableness (IPIP-2) is low and both conscientiousness (IPIP-3) and intellect (IPIP-5) are high or

low, probabilities for both proposal and acknowledgement are high. When intellect (IPIP-3) is low, the probabilities of communication shift slightly to acknowledgement. Also, high agreeableness (IPIP-2) and conscientiousness (IPIP-3) appear to promote acknowledgement responses.

Results of probabilities of posts transitions for IL are summarised in Table 8. Any combination of levels for IL do not appear to affect the frequencies of proposal, though the ability of IL is expected to contribute to participant's level of communication. Combinations of both 'high-high' (high IL-1 and high IL-2) and 'low-low' (low IL-1 and low IL-2) for two factors promote the probability of acknowledgement. A combination of 'high-low' (high IL-1 and low IL-2) also affect the frequency of acknowledgement.

Table 8 Probabilities for communication actions using IL

<i>Skill</i>	<i>Attitude</i>	<i>Pro</i>	<i>Per</i>	<i>Enc</i>	<i>Ack</i>
High	High	0.28	0.13	0.12	0.46
Low	High	0.25	0.25	0.25	0.25
High	Low	0.25	0.20	0.22	0.33
Low	Low	0.25	0.13	0.18	0.43

Notes: Pro: proposal, Per: permission, Enc: encouragement, and Ack: acknowledgement.

These results suggest that factors related to a participant's characteristics contribute to the appearance of functional communication in the discussion.

5 Discussion

The game products created as a result of team collaborations in this study were evaluated using a system usability scale with scores assigned by participants from another team. The mean scores were calculated for every cycle and for every team. The relationships between system usability scores [0–100] and categories of communication action were analysed. The results suggest that the frequency of proposal contributed to usability scores during the first cycle. Constructive communication in the early stages positively affected the resultant products (Teranishi et al., 2017a).

Some significant relationships between scores of participant's reflections and frequencies of communication activities were examined in Table 6. Most contributions of communications are observed during the second and third cycles. Therefore, participant's reflections appear to be affected by communication frequencies in the latter two cycles, especially the appearance of proposal communications.

These results suggest different levels of factor scores of effectiveness of communication activity. The frequency of constructive communication, such as proposals in the early stages, affected the quality of the product, and the frequency of constructive communication in the latter stages positively influenced participant's reflections. These points can be used to support the efficacy of collaborative communications.

Team communication may depend on a combination of participant's characteristics. The results in Tables 7 and 8, suggest that participant's characteristics affect the frequency of communication activities. Some combinations promote the frequency of proposal communications. Though the communication category acknowledgement is

the most frequent category of activity, its frequency is influenced by combinations of participant's characteristics. When participant's levels of factor scores of characteristics are low, the probability of acknowledgement communication is high. Acknowledgement communication can be easily produced in the form of a sympathetic response. Two types of relationships between participant's characteristics and their probabilities for acknowledgement communication can be explained as positive agreement and sympathising. To activate collaboration communication, a combination of characteristics of participating team members should be considered, such as team regard to their characteristics, and providing information to the facilitators for team collaboration.

As this study focused mainly on learning progress, most analysis was applied to learning activities across three cycles of development, and the learning progress throughout the collaborative work as discussed. Since the experiment is based on a collaborative learning project, learning performance of participants was evaluated objectively. The project was initially designed as an open-ended learning exercise over three cycles. Hence, the observed measures of individual participants were assessed at the end of the third cycle. A more thorough evaluation is the subject of future research. The results and discussions reported here are based on analysis of a small number of participants as a case study. In particular, for all participants recruited for this experiment, some motivation factors might affect their behaviours, such as the Hawthorn effect, as well as team learning conditions – the number of groups and team size (Shaw, 2013). This is despite the three sequential sessions being conducted on all participants, and all collaborating teams successfully improving continuously. To confirm the validity and dependencies for the contents of collaboration in this study, a further study with more participants is needed. It is hoped these points will be determined in future experiments by introducing some more comparative evaluations.

6 Conclusions

In this study, the contributing factors to the usability scores in a game development activity were investigated using cross-cultural TBL. Based on the usability scores, the relationships with the learners' reflections, communications and attributes were analysed. The frequency of appearance of communication in four categories was also analysed to evaluate the effectiveness of online communication using social networking tools to enhance collaborative work.

The following results are reported.

- 1 The examination between SUS scores and the participants' reflection indicates that the usability scores decreased when the participants felt difficulties in their collaboration. In order to yield a better result, teamwork is the most significant factor in TBL.
- 2 The frequency of appearance of communication in four categories was analysed, such as proposal, permission, encouragement, and acknowledgement. The four types of communication activities were defined, and participant's posts classified, according to the type of communication.
- 3 Based on the relationship between SUS scores and communications, proposal and acknowledgement, communications contribute to the usability scores in the earliest

stage of the development cycles, in order to convey a better usability result. Hence, we should encourage students to have these types of communications within and between teams for better discussions and team activities overall.

- 4 According to the relationship between the SUS scores and the participants' attributes, information skills were related to the implementation performance as well as SUS scores.
- 5 Some significant relationships between participant's reflections and the frequency of the four types of communication were recognised. This suggests that the frequencies in the latter cycles contributed significantly to scores which were based on reflections.
- 6 The transitions of appearance of the four types of posted messages were summarised as a Markov model. The appearance of probabilities which depend on team members' characteristics were also calculated using a Bayesian network technique. The results show that some scores for personality and IL of participants influenced the probability of messages being produced.

In regards to these results, development of appropriate support procedures should be considered for online work requiring team collaboration, and will be the subject of further study.

Acknowledgements

This work was partially supported by an International Collaboration Research Promotion Fund, School of Engineering, Tokyo Institute of Technology. Also, KAKEN (17H00825) partially supported this work.

This paper is an extended version which is based on reports at ECSM2017 (Teranishi et al., 2017b), IV2017 (Teranishi et al., 2017a), and SAC2018 (Teranishi et al., 2018). The authors would like to thank those who provided useful comments regarding these papers.

References

- Aleman, C.G. and Vangelist, A.L. (1994) 'Protocol analysis as a method for analyzing conversational data', in *Proc. of the Annual Meeting of the Western States Communication Association*, pp.1–17.
- Armstrong, J. and Hyslop-Margison, E. (2006) 'Collaborative learning and dialogue: democratic learning in adult education', *New Horizons in Adult Education and Human Resource Development*, Vol. 20, No. 4, pp.6–15.
- Bhuasiri, W., Xaymoungkhoun, O., Zo, H., Rho, J.J. and Ciganek, A.P. (2012) 'Critical success factors for e-learning in developing countries: a comparative analysis between ICT experts and faculty', *Computers & Education*, Vol. 58, No. 2, pp.843–855.
- Bottcher, S.G. and Dethlefsen, C. (2015) *R Packages 'Deal'* [online] <http://cran.r-project.org/web/packages/deal/deal.pdf> (accessed 31 July 2017).
- Bradford, G. and Wyatt, S. (2010) 'Online learning and student satisfaction: academic standing, ethnicity and their influence on facilitated learning, engagement, and information fluency', *Internet and Higher Education*, June, Vol. 13, No. 3, pp.108–114.

- Brooke, J. (1996) 'Sus-a quick and dirty usability scale', *Usability Evaluation in Industry*, Vol. 189, No. 194, pp.4–7.
- Campos, A.A.M. and ao Correia de Freitas, J. (2015) 'Assessing and implementing English-learning mobile applications in a university graduation program: SLA 2.0', in *Proceedings of CSEDU 2015 – Doctoral Consortium*, pp.8–15.
- Chen, S.-J. and Caropreso, E.J. (2004) 'Influence of personality on online discussion', *Journal of Interactive Online Learning*, Vol. 3, No. 2, pp.1–17.
- Chen, T.L. (2014) 'Exploring e-learning effectiveness perceptions of local government staff based on the diffusion of innovations model', *Administration & Society*, Vol. 46, No. 4, pp.450–466.
- Cheng, G. and Chau, J. (2016) 'Exploring the relationships between learning styles, online participation, learning achievement and course satisfaction: an empirical study of a blended learning course', *British Journal of Educational Technology*, Vol. 47, No. 2, pp.257–278.
- Cheston, C.C., Flickinger, T.E. and Chisolm, M.S. (2013) 'Social media use in medical education: a systematic review', *Academic Medicine*, Vol. 88, No. 6, pp.893–901.
- Chi, M.T. and Menekse, M. (2015) 'Dialogue patterns that promote learning', in Resnick, L., Asterhan, C. and Clarke, S. (Eds.): *Socializing Intelligence through Academic Talk and Dialogue*, Chapter 21, pp.263–274, AERA, Washington, DC.
- Combéis, S., Beresnevičius, G. and Dagienė, V. (2016) 'Learning programming through games and contests: overview, characterisation and discussion', *Olympiads in Informatics*, Vol. 10, pp.39–60.
- Daly, J.A., Weber, D.J., Vangelisti, A.L., Maxwell, M. and Neel, H. (1989) 'Concurrent cognitions during conversations: protocol analysis as a means of exploring conversations', *Discourse Processes*, Vol. 12, No. 2, pp.227–244.
- Echeverría, L., Cobos, R., Machuca, L. and Claros, I. (2017) 'Using collaborative learning scenarios to teach programming to non-CS majors', *Computer Application in Engineering Education*, September, Vol. 25, No. 5, pp.719–731.
- Erkens, G. and Janssen, J. (2008) 'Automatic coding of dialogue acts in collaboration protocols', *International Journal of Computer-Supported Collaborative Learning*, Vol. 3, No. 4, pp.447–470.
- Fujii, Y. (2007) 'Development of a scale to evaluate the information literacy level of young people – comparison of junior high school students in Japan and Northern Europe', *Japan Journal of Educational Technology*, Vol. 30, No. 4, pp.387–395.
- Gatignon, H. (2014) 'Cluster analysis', in Gatignon, H. (Ed.): *Statistical Analysis of Management Data*, Chapter 12, pp.453–285, Springer, Boston, MA.
- Goldberg, L. (1999) 'A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models', *Personality Psychology in Europe*, Vol. 7, pp.7–28.
- Gomez, E.A., Wu, D. and Passerini, K. (2010) 'Computer-supported team-based learning: the impact of motivation, enjoyment and team contributions on learning outcomes', *Computers & Education*, Vol. 55, No. 1, pp.378–390.
- Hara, N., Bonk, C.J. and Angeli, C. (2000) 'Content analysis of online discussion in an applied educational psychology course', *Instructional Science*, Vol. 28, No. 2, pp.115–152.
- Hill, P. (2012) 'Online educational delivery models: a descriptive view', *EDUCAUSE Review*, Vol. 47, No. 6, pp.84–97.
- Hrastinski, S. (2009) 'A theory of online learning as online participation', *Computer & Education*, January, Vol. 52, No. 1, pp.78–82.
- International Personality Item Pool (IPIP) (2004) 'A scientific collaboratory for the development of advanced measures of personality traits and other individual differences' [online] <http://ipip.ori.org>.
- Jung, I., Choi, S., Lim, C. and Leem, J. (2002) 'Effects of different types of interaction on learning achievement, satisfaction and participation in web-based instruction', *Innovations in Education and Teaching International*, Vol. 39, No. 2, pp.153–162.

- Kanbul, S. and Uzunboylu, H. (2017) 'Importance of coding education and robotic applications for achieving 21st-century skills in North Cyprus', *International Journal of Emerging Technologies in Learning*, Vol. 12, No. 1, pp.130–141.
- Kebritchi, M., Lipschuetz, A. and Santiago, L. (2016) 'Issues and challenges for teaching successful online courses in higher education: a literature review', *Journal of Educational Technology Systems*, Vol. 46, No. 1, pp.4–29.
- Klemmer, S. (2016) *Human-Computer Interaction (MOOC)* [online] <https://www.coursera.org/lecture/human-computer-interaction/human-computer-interaction-pE6EB> (accessed 27 February 2015).
- Lye, S.Y. and Koh, J.H.L. (2014) 'Review on teaching and learning of computational thinking through programming: what is next for K-12?', *Computers in Human Behavior*, Vol. 41, pp.51–61.
- Manca, S. and Ranieri, M. (2017) 'Implications of social network sites for teaching and learning. Where we are and where we want to go', *Education and Information Technologies*, Vol. 22, No. 2, pp.605–622.
- Marra, R.M., Moore, J.L. and Klimczak, A.K. (2004) 'Content analysis of online discussion forums: a comparative analysis of protocols', *Educational Technology Research and Development*, Vol. 52, No. 2, p.23.
- Menekse, M. and Chi, M.T. (2019) 'The role of collaborative interactions versus individual construction on students' learning of engineering concepts', *European Journal of Engineering Education*, Vol. 44, No. 5. pp.702–725.
- Michaelsen, L.K., Knight, A.B. and Fink, L.D. (2002) *Team-Based Learning: A Transformative Use of Small Groups*, Greenwood Publishing Group, Santa Barbara, CA, USA.
- Moore, M.G. and Kearsley, G. (2011) *Distance Education: A Systems vView of Online Learning*, Cengage Learning, Boston, MA, USA.
- Nakayama, M., Yamamoto, H. and Santiago, R. (2008) 'Impact of information literacy and learner characteristics on learning behavior of Japanese students in online courses', *International Journal of Case Method Research & Application*, Vol. 20, No. 4, pp.403–415.
- Nakayama, M., Mutsuura, K. and Yamamoto, H. (2015) 'Analysis of student's reflections upon the efficacy of their note-taking activities in a blended learning course', in *2015 International Conference on Information Technology Based Higher Education and Training (ITHET)*, IEEE, pp.1–6.
- Ogot, M. and Okudan, G.E. (2006) 'The five-factor model personality assessment for improved student design team performance', *European Journal of Engineering Education*, Vol. 31, No. 5, pp.517–529.
- Open University (2014) *Innovating Pedagogy 2014* [online] <http://www.open.ac.uk/innovating> (accessed 2 February 2016).
- Paechter, M., Maier, B. and Macher, D. (2010) 'Students' expectations of, and experiences in e-learning: their relation to learning achievements and course satisfaction', *Computers & Education*, Vol. 54, No. 1, pp.222–229.
- Pintrich, P.L. and Groot, E.V.D. (1990) 'Motivation and self-regulated learning components of classroom academic performance', *Journal of Educational Psychology*, Vol. 82, No. 1, pp.33–40.
- Prasolova-Førland, E., Wyeld, T. and Chang, T-W. (2007) 'Constructing a virtual tower of babel: a case study in cross-cultural collaboration across three continents', in *International Conference on Virtual Systems and Multimedia*, Springer, pp.143–153.
- Prasolova-Førland, E., Steinsbekk, A. and Lindseth, F. (2017) 'Practicing interprofessional team communication and collaboration in a smart virtual university hospital', in *International Conference on Smart Education and Smart E-Learning*, Springer, pp.191–224.
- Rasiah, R.R.V. (2014) 'Transformative higher education teaching and learning: using social media in a team-based learning environment', *Procedia-Social and Behavioral Sciences*, Vol. 123, pp.369–379.

- Robins, A., Rountree, J. and Rountree, N. (2003) 'Learning and teaching programming: a review and discussion', *Computer Science Education*, Vol. 13, No. 2, pp.137–172.
- Seaton, D.T., Bergner, Y., Chuang, I., Mitroa, P. and Pritchard, D.E. (2014) 'Who does what in a massive open online course?', *Communication of the ACM*, Vol. 57, No. 4, pp.58–65.
- Shaw, R.S. (2013) 'The relationships among group size, participation, and performance of programming language learning supported with online forums', *Computers & Education*, Vol. 62, pp.196–207.
- Teranishi, A., Nakayama, M., Wyeld, T. and Eid, M. (2017a) 'Analysis of game development activity using team-based learning', in *Proceedings of 21st International Conference Information Visualisation*, pp.211–217.
- Teranishi, A., Nakayama, M., Wyeld, T. and Eid, M. (2017b) 'Effectiveness of social media communication in game development study using team-based learning', in *Proceedings of 4th European Conference on Social Media*, pp.289–296.
- Teranishi, A., Nakayama, M., Wyeld, T. and Eid, M. (2018) 'Online team-based game development discussions pattern summarised using probabilistic models', in *Proceedings of the 33rd Annual ACM Symposium on Applied Computing (SAC)*, pp.234–239.
- Traphagan, T., Kucsera, J.V. and Kishi, K. (2010) 'Impact of class lecture webcasting on attendance and learning', *Educational Technology Research & Development*, February, Vol. 58, No. 1, pp.19–37.
- Vahldick, A., Mendes, A.J. and Mendes, M.J. (2014) 'A review of games designed to improve introductory computer programming competencies', in *Proceedings of 2014 IEEE Frontiers in Education Conference (FIE)*, pp.1–7.
- Wu, B. and Wang, A.I. (2012) 'A guideline for game development-based learning: a literature review', *International Journal of Computer Games Technology*, Vol. 2012, pp.1–21.
- Wyeld, T. and Barbuto, Z. (2014) 'Don't hide the code!: empowering novice and beginner programmers using a HTML game editor', in *Proceedings of 18th International Conference Information Visualisation*, pp.125–131.
- Wyeld, T. and Nakayama, M. (2018) 'A study in introducing adolescent students to simple programming skills using an online games editor', in *Proceedings of International Conference on Technology in Education ICTE2018*, pp.181–198.
- Wyeld, T. (2010) 'Working with cultural differences: a case study in multicultural teamwork using a 3dave', *Educational Games: Design, Learning and Applications*, pp.137–149, Nova Science Pub Inc., Bombay.