Comparing User Experience and Performance in SecondLife and Blackboard

Sutcliffe, A.G. & Alrayes, A.
Manchester Business School, University of Manchester
Booth Street West, Manchester M15 6PB, UK
ags@man.ac.uk

Abstract. Collaborative problem solving was compared in SecondLife (SL) and Blackboard (BB) and both technologies were compared with a face-to-face (FTF) control condition. There were no performance differences overall, although FTF was quicker and preferred, followed by BB and SL. BB was perceived to be more usable, whereas SL provided better user experience. Worse performance was indicated by dislike of avatar interaction in SL, and poor user experience in BB, whereas better performance was associated with engagement with avatars, and better usability in BB. The affordances for collaboration in each technology are discussed, with reflections on the mixed methods approach using qualitative and quantitative data analysis.

Keywords. Collaborative problem solving and learning, Affordances, Mixed methods evaluation.

1 Introduction

The shortcomings of collaborative virtual environments (CVEs), such as a poor sense of presence, and limited non-verbal communication, have been pointed out by several authors [11]. However, CVEs are clearly successful for multiplayer games (such as World of Warcraft [5]), which has led educators to explore the potential of Second Life as a CVE that might motivate collaborative learning [9,22]. Some studies of collaboration in CVEs have suggested that presence and user experience can be superior to conventional 2D interfaces since avatars provide improved awareness of others and shared tasks [23,31]. However, comparisons of performance between CVEs and real-world equivalents have failed to show any clear advantage for virtual worlds. For example, in training medical students in interviewing skills for diagnosis, the VE performed as well as real patients but was less satisfying [13]. SL can facilitate sharing experience and personal information, although Neustaedter et al.'s [17] qualitative study suggests that SL complements real-life experiences rather than being an effective substitute for them.

Several issues and criteria have been proposed in Computer Supported Collaborative Work (CSCW) evaluation frameworks [21], such as group characteristics, situation factors (context), individual characteristics, task properties, group process, and task and group outcomes. Neale et al. [16] added shared awareness to this list; however, as they pointed out, shared awareness is complex and involves knowledge of others' roles, identities, goals, and activities. Comparing affordances for collaboration in 3D virtual environments and conventional 2D interfaces was an initial motivation for this study. Evaluations in collaborative technology have tended

to follow either an ethnographic approach to investigate the context of use in depth, or more focused experimental analyses directed towards specific questions about collaboration (e.g. [19]). A second motivation for this study was to investigate the mixed methods approach for understanding how affordances for collaboration contribute to user experience as well as performance and learning.

In this paper we investigate interaction and collaboration in SecondLife (SL) as an example of a 3D desktop CVE and Blackboard (BB) as a more traditional collaborative interface. The study did not examine collaborative learning per se, where longer-term investigation would be necessary; instead, our focus was on understanding how user experience and motivation may create sufficiency conditions for collaborative learning in one Computer Supported Collaborative Learning (CSCL) technology (BB) and one potentially more engaging technology (SL) which is being adapted for CSCL purposes. Two research questions are addressed: first, will the motivation of interacting in SL improve performance and user engagement in comparison to BB? Performance is this case was effectiveness of collaborative problem solving. The second question concerned investigating the quality of the user experience and possible influences of experience on performance with different technologies, using a mixed (qualitative and quantitative) methods approach. In following sections of this paper, related work in CVEs and CSCL technologies is reviewed, then the experimental methods are described. This is followed by quantitative data analysis and qualitative investigations into user experience.

2 Related Work

Studies of CSCL environments have reported improved levels of learning [8] and participation [7,12] compared to a face-to-face (FTF) interaction. In contrast, other studies have found that students working in CSCL environments can perceive discussions as more confusing [28] and less productive [25]. Furthermore, CSCL environments can produce lower levels of student participation [13], higher levels of conflict [10], poor group cohesion [25] and lower levels of satisfaction [1]. These contradictions suggest further in-depth studies of the causes of either failure or success of CSCL technology are necessary.

Familiarity among participants is well known to have a beneficial effect on technology mediated group work [19], since knowledge of others can enhance social awareness and organisation of work. However, in CSCL previous research is divided about the effect of group members' familiarity and performance. Newcomb and Bagwell [18] found collaboration between friends produced more intense social activity, more frequent conflict resolution, more effective task performance, greater equality, and loyalty. Familiar groups may also help creation of effective shared problem-solving spaces [2,3].

In contrast, Maldonado et al. [15] found a strong negative correlation between prior friendship and groupwork performance. Furthermore, working with friends can yield lower-quality outcomes because friends indulge in more non-task related behaviour, with stronger pressures to agree, and unwillingness to be critical of each others' ideas [6,34]. Groups of friends may also disagree more frequently [24] and

become more concerned with resolving disagreements [18]. Thus it appears the jury is out on the effect of social relationships in collaborative learning.

User experience (UX) has many definitions arising from research on aesthetics [29,30] which extended traditional concepts to show that users' response to interaction has an affective dimension. However, aesthetics alone fails to capture the interactive elements of user experience especially in virtual worlds such as SL [27], so we adapt concepts from virtual reality [31,33] to augment measures of UX.

3 Methods

The hypothesis that interactive 3D worlds would motivate more effective collaboration and learning was investigated by comparing cooperative problem solving in SL and BB, a leading 2D collaborative learning environment. The role of affordances for technology-mediated collaboration were investigated to evaluate how 2D user interface features in BB compared to 3D interface in SL.

Sixty three participants (30 male, 33 female) took part in the experiment, organised in 21 teams with 3 members in each team. The participants were 24 undergraduate students, 8 postgraduate (Masters) students and 31 postgraduate (PhD) students at Manchester University. The participants were allowed to select their own team members. The majority (15 out of 21) of the teams were mixed gender. All participants were familiar with BB so no further training was necessary. Only one had used SL so training in user interface operation, avatar control and communication was given.

The teams had to solve three analogous versions of a survival prioritisation problem. Three different analogues were used (Lost in the Desert, Lost at Sea and Survival on the Moon) to minimise direct learning effects across repeated trials. Participants were presented with lists of items which might be useful or useless for survival in each scenario environment. The goal in each scenario was to collaboratively decide the best ranking of the items in order to survive. The items and context differed so no direct transfer of problem solutions was possible, although meta-level and process learning, i.e. approach to the problem, and organising collaborative problem solving was possible. The prime hypothesis therefore focused on performance:

Since motivation and user engagement with CSCL technologies are important precursors for effective learning, the secondary motivation for the study was to investigate the differences in interaction and collaboration between the two technologies (SL and BB) by comparing them with the control FTF condition.

A within-groups repeated measures design compared BB, a standard non-graphical user interface and SL, a 3D graphical environment with avatar-mediated interaction, with a FTF control condition. The order of tasks and technology conditions was counterbalanced. The independent variables were modality conditions and collaborative functions provided by the technology:

FTF: all modalities plus co-presence. Support for collaboration: notepad for handwritten records.

BB: asynchronous text message, e-mail and chat. Support: message threading, exchange of documents as attachments, electronic notes created using Word.

SL:　visual communication via avatars, movement, position and limited gesture; near-synchronous text chat. Support: manipulation of text objects and lists in the virtual world.

In the control FTF condition the three team members were co-present in one room so they could converse naturally to arrive at a solution. They had access to the problem narrative and notepads to make lists or notes. The solution was recorded as a word processed list of objects in rank order. In BB each team member worked on a separate PC connected to BB (Figure 1a). The solution was presented as a word processed list of items ranked in order of their utility for survival. In SL each team member worked on a separate PC and was assigned an avatar of their own gender. They could move, pick up and manipulate numbered objects which could be placed in slots by item on a text list to denote the rank order. Hence the solution could be collaboratively constructed and was presented as a ranked list created in the 3D environment. Communication was by text chat bubbles associated with each avatar (see Figure 1b).
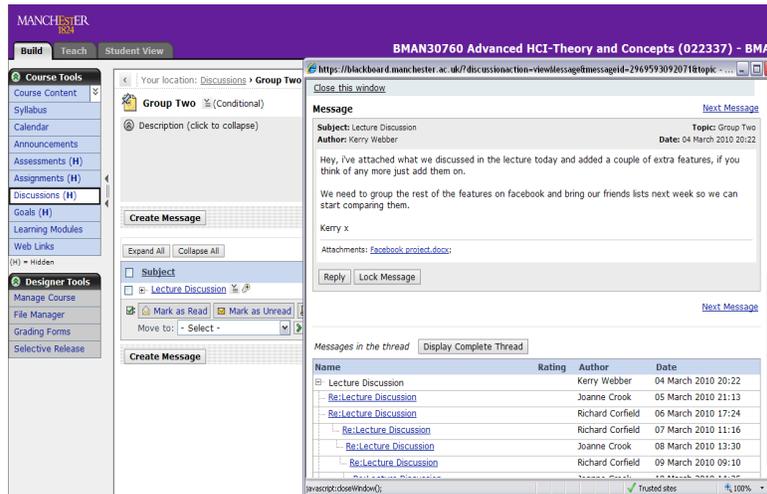

Figure 1a. Blackboard user interface showing a group discussion thread
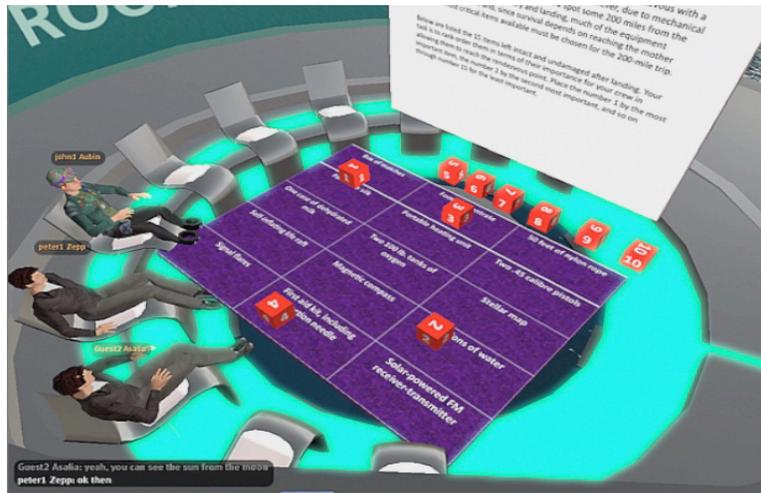
Figure 1b. Screen dump showing SecondLife with group discussion

Voice communication was not used, since we wished to focus on the avatar versus conventional UI interaction, rather than compare communication modalities. Furthermore text chat is commonly used in SL when audio output can be annoying in public contexts such as co located class with several groups. Text communication had the added advantage of recorded transcripts of conversations for qualitative analysis. No verbal or FTF visual contact was allowed in either the BB or SL conditions.

After completing the task, each participant filled in a questionnaire with four sections rating user experience on 1-7 Likert scales (affect, general measures of UX, interaction quality, and presence). A further questionnaire rated satisfaction and perception of their own performance (decisions made, challenge, motivation) and overall preference. The experiment ended with a team interview to investigate the participants' experience, feelings and reflections on collaboration using the two technologies and in the control conditions. Questions probed the participants' perceptions of the quality of interaction, interactive experience with the technologies, critical incidents and usability problems with reflections on user experience in each condition. The participants' behaviours were observed during the experiment by taking notes and video recordings in SL sessions. BB chat room logs were saved for further analysis of users' discourse. Observations focused on participants' interaction with the technologies, artefacts in the real world or representations in BB and SL, and patterns of communication within the groups.

To summarise, the dependent variables were: group performance (task completion times, errors) individual participants' ratings of the quality of interaction and collaboration with each technology and FTF. The post-test rating data was ranked by means within each condition to simply presentation of the results. Performance differences between the technologies and scenario tasks were tested by ANOVAs, and post hoc tests. Qualitative data analysis focused on patterns of collaboration (observation and video logs) and the participants' recollection of the effectiveness of group working, critical incidents and breakdowns. Interview transcripts were coded

following the open coding and axial/categorical aggregation conventions [26], then the transcripts were inspected for excerpts which illustrated the more frequent topics and issues reported by the respondents.

4   Results- Quantitative Data

Social closeness was calculated by summing the relationship strength reported by each group member on an 8-point scale and calculating the percentage of the maximum possible score (48). None of the groups contained three close friends (M 23.7%, range 12-65%) and only five groups reported an aggregate closeness >50%. No correlations were found between performance and social closeness in the groups. Indices for group interaction were calculated from the overall number of contributions factored by the contribution ratio:

Interaction = TotContrib * (highest/lowest member contributions)

There were no correlations between these measures and performance accuracy and times, so the quantity of interaction within the groups did not appear to influence performance overall or in any of the conditions.

   Not surprisingly, since BB was the university's standard collaborative learning environment, most participants had greater experience with BB (M 4.91) than with SL (M 2.13 on a 7-point scale, range 1=once to 7=daily use). The participants had limited knowledge of the scenarios (M 2.91) so for most it was a novel problem.

Performance

Expert solutions for each problem analogue were taken from the literature and used to calculate the group performance by scoring goodness-of-fit for the priority order of the ideal solutions against each group's solution. There was a significant difference between the task scenarios $F_{(2,8)}=74.043$, $p<0.000$, with solutions for the Moon survival scenario being more accurate than the other two, but there was no main effect on performances between the three environments, and no interaction; see Table 1.

Table 1. Percentage of correct scores by scenario and technology

	Desert	Sea	Moon
FTF	47	43	70
SL	49	47	74
BB	53	49	76

   All groups completed the task successfully although the average scores in the Desert and Sea scenarios showed room for considerable improvement; see Table 1. There were no apparent learning effects and the groups did not improve their performance after successive exposures to the task.

   Groups completed the task more rapidly FTF, as might be expected (M 11.23 minutes), compared to 26.80 mins for SL and 24.00 mins for BB; while the Desert scenario was completed more quickly (M 18.47) than the Moon (M 21.53) or Sea scenarios (M 22.03). There was a significant main effect for technology (F=13.5 (2,28) p<0.001) but not for task. FTF was faster than the technologies but there were

no differences between SL and BB (post hoc tests p<0.05). The Moon task was completed more accurately although it took slightly longer than the Desert scenario. In the technology conditions, FTF was quicker, but BB was more accurate, although these difference were small and not significant.

Group interaction
Interaction within the groups, measured by message exchange and chat posting, was more frequent in BB (M 113.87 messages), than SL (M 107.40). There were slightly fewer interactions for the Moon task (M 103.55) than for Desert (M 106.20) or Sea (M 122.15). Although there were no significant main effects, the interaction (scenario x technology) was marginally significant F(2,9) 3.48, p<0.05, with more interactions in SL for the Sea and Moon tasks and more in BB for the Desert task. FTF interactions were not recorded since dialogue exchanges were too rapid for accurate recording, so no direct comparison was possible; however, informal observation suggested that FTF interaction was most frequent.

Post-test questionnaire ratings
There were significant differences in nearly all the post-test rating scores, with the control condition tending to be most favoured, followed by BB with SL in third position (p>0.001 ANOVAs, with post hoc tests on individual ratings). The results of post hoc tests are reported in rank order of mean scores where, by default, differences between all three combinations were at least p<0.05, apart from the table cells shaded to show where the differences between the technologies were not significant even though they did differ from FTF. In the affect measures, illustrated in Table 2, FTF was first for pleasure and joy, while SL was first for both the positive emotion of surprise and the negative emotions of anxiety, frustration, fear and disgust. In all categories except pleasure, BB had lower mean scores than SL, but these differences were not significant.

Table 2. Rank order of means of the affect rating scales.
shading = no significant difference between technologies (BB and SL)

Measure	Rank order 1-2-3 by means		
Pleasure	FTF	BB	SL
Surprise	SL	BB	FTF
Anxiety	SL	BB	FTF
Joy	FTF	SL	BB
Frustration	SL	BB	FTF
Disgust	SL	BB	FTF
Fear	SL	BB	FTF

In the user experience measures illustrated in Table 3, not surprisingly FTF was most positive; however, SL had higher means than BB for all measures except the motivation for repeated use (use again), although these differences were not significant.

Table 3. User experience rank order by means

Hold your attention	FTF	SL	BB
Feel excited	FTF	SL	BB
Good mood after using	FTF	SL	BB
Use again	FTF	BB	SL
Vivid memory	FTF	SL	BB
Memory good or bad	FTF	SL	BB

The FTF control condition was most favoured, then BB was a clear winner over SL in all measures, with a significant difference for clear design (p>0.05 post hoc test); see Table 4.

Table 4. Interaction quality measures, rank order by means

Convenient use	FTF	BB	SL
Easy to use	FTF	BB	SL
Easy to navigate	FTF	BB	SL
Clear design	FTF	BB	SL

Interaction quality measures indicated that SL was perceived as complex and challenging than BB (p>0.05) while BB gave the best awareness of external events, possibly reflecting the information-intensive user interface. SL was ranked second to FTF in engagement and (less) awareness of the user interface, realism indicating some benefit from presence and immersion in the 3D graphical world and avatars; but BB was rated better for natural feel than SL (p<0.05) and being absorbed, so the usability problems in SL (Table 5) may have disturbed the sense of presence.

Table 5. Ranks order by means for interaction quality and presence from [33]
Shading = no significant difference between technologies (BB and SL)

Good pace (speed) of interaction	FTF	BB	SL
Pace of interaction interesting	FTF	BB	SL
How complex was the interaction	SL	BB	FTF
How challenging	SL	BB	FTF
How engaged	FTF	SL	BB
Awareness of user interface	FTF	SL	BB
Natural feel	FTF	BB	SL
Awareness external events	BB	SL	FTF
Totally absorbed	FTF	BB	SL
Realistic	FTF	SL	BB

As expected, overall satisfaction favoured FTF; (see Table 6); F(2,267)=45.88, p<0.000); while participants were more satisfied with BB than SL, although this difference was only marginally significant (p<0.05 in post hoc tests).

Table 6. Environment satisfaction

Environment	Mean	SD
FTF	6.14	1.148
BB	4.90	1.41
SL	4.35	1.504

Satisfaction with decisions made followed the same pattern, illustrated in Table 7 $(F(2,267)=10.725, p<0.001)$ with BB in second place; the difference between BB and SL was not significant, although FTF was better than both BB and SL $(p<0.01,$ post hoc tests).

Table 7. Satisfaction with decisions made

Environment	Mean	SD
FTF	6.17	1.199
BB	5.50	1.105
SL	5.13	1.561

SL was considered to be more of a challenge than the other conditions (Table 8), probably because it was less familiar to the participants $(F(2,267)=20.469, p=<0.0001)$, while FTF was best for overall motivation $(F(2,267)=20.469, p<0.001)$ with BB in second place and SL third, although the means were not significantly different.

Table 8. Challenge and overall motivation scores

		Mean	SD
Challenge	FTF	3.31	2.053
	SL	4.71	1.637
	BB	4.16	1.476
Overall motivation	FTF	5.93	1.428
	SL	4.60	1.654
	BB	4.81	1.413

To summarise, performance was no better FTF than with SL and BB technologies, which is surprising since the CSCW technologies can constrain discussion and fluid interaction. The control FTF was rated best on experience, operational ease of use, positive affect and overall satisfaction. SL was in second position for user experience items, whereas BB was second on usability measures, although most differences were not significant. SL evoked mixed emotions with both surprise and fear, anxiety and frustration. For overall satisfaction there no significant difference between SL and BB.

5   Results: Qualitative Data Analysis

The patterns of positive/negative comments in the post-test interviews agreed with the quantitative measures, demonstrating that most individuals preferred FTF interaction over SL or BB, citing natural communication as the main reason (30%), followed by ease of interaction (22%) and rapid interaction (18%). An equal number of positive and negative comments (44) were reported for BB, with familiarity (20%) and ease of use (19%) heading the list, followed by simple interface, easy to focus and communication (11%). The more frequent negative comments were boring interaction (30%), poor response time (25%), didn't like text-only communication (18%), and hindered discussion (14%).

SL attracted more negative than positive comments overall (55%). Among the more frequent positive aspects were enjoyable and fun (41%), close to reality (34%) and interaction quality (15%). In contrast, the downsides were unfamiliarity (14%), distracting interface (14%), and navigation problems (15%), with most other issues relating to complexity, general frustration and usability issues.

The general themes which emerged for BB reflected users' perceptions of a simple and familiar interface which was effective for the task in hand, although it was boring and not very exciting. The collaborative affordances were limited to communication and document exchange. In contrast, SL was perceived to be more dynamic, stimulating and interesting, but with downsides of being distracting and annoying. Collaboration afforded by avatars was motivating in terms of curiosity but not seen as relevant to the task. Several comments were made that the avatars were not faithful representations and the absence of facial expressions reduced the sense of presence. Furthermore, the limited gestures and difficulties in moving them reduced the effectiveness of avatars, with several respondents commenting that interaction evolved into a chat room format making the avatars irrelevant. For example:

> SL is very restrictive, as not only do you have to interact with a keyboard, you also have to use a mouse to interact and move around. The avatars are a distraction, because even though they represent you, they do not show emotion or body language. Using the mouse to interact with items is hard.

> SL provided too many distractions. For example, moving around and moving boxes made it harder to make a good choice. BB was a far simpler interface which made it easier to focus on the task.

Comments in favour of SL usually focused on the user experience and novelty:

> I really enjoyed the SL experience. Maybe as it is my first time so it's quite novel. I also felt hugely immersed in this environment in comparison to BB. BB wasn't much fun. SL was more absorbing due to the avatars and looking at how they behaved in SL.

And one of the few who did not rate FTF first commented:

I prefer using SL and BB over FTF because you get in contact with different types of people without the need to get involved with them in real life. And SL is more fun, interesting and less boring that FTF.

Pointing to and manipulating the numbered objects in SL was infrequent, and most groups discussed the prioritisation order; one individual was nominated to move the object to record the result.

Discussion was effective in all groups; however, individuals in four groups reported excluding or ignoring the opinions of others, and six groups commented that, given the opportunity to repeat the task, they would aim to be more inclusive and improve discussion.

The quantitative data suggested that overall there was little to choose between BB and SL; in contrast, the qualitative data analysis suggested different reasons for liking or disliking each technology. The interviews were analysed to see if the performance in the five best and five worst groups showed any patterns that might link users' reactions to the technologies and their performance.

Comparison of the Top 5 and Bottom 5 groups
Performance among the groups was normally distributed, so groups in the tails of the distribution were investigated in more depth.

Comments made by the five best and five worst performing groups were compared, to tease apart possible reasons for success and failure between the conditions. In FTF, the comments of the top and bottom five groups were nearly all positive and followed the general pattern of favouring natural communication ability to see facial expressions and non-verbal communication, use speech efficiently, and the social advantages of being able to get to know other group members. The best performing groups made more comments overall and more positive (60%) than negative comments; conversely, the five worst performing groups made more negative (57%) than positive comments. The distribution of comments by valency is shown in Table 9.

Table 9. Valency of comments expressed as percentage of total comments separately for the top 5 and bottom 5 groups

	Best 5 groups		Worst 5 groups	
	+ve	-ve	+ve	-ve
SL	30	15	22	35
BB	30	35	10	33

For the five best groups, positive comments were evenly distributed, but they made more negative comments about BB. However, most of these comments were criticisms about the need for improving its functionality. The five worst groups made a similar number of negative comments about both technologies, with most comments concerning poor usability. These groups were more positive about SL, with comments on the engagement of the 3D world and avatars. So, even though both groups made a similar number of negative comments about BB, for the better performers these were

complaints about missing requirements whereas the worst performers were reporting usability problems. This tendency was also apparent for SL.

For BB, the five top groups reported that communication was quick, easy and efficient, with downsides of not being exciting and text-only communication being limited. In contrast, the five bottom groups reported more negative comments such as "boring interface, limitation of text-only interaction, slow, frustrating and difficult for effective discussion".

In SL, the reasons were quite different. The better performing groups all made favourable comments about the sense of presence, interesting and exciting interaction, reporting novel and exciting experience with few negative comments. In contrast, the poorly performing groups all made frequent criticism that SL was difficult to use and navigate, the graphics and interaction felt artificial, also that it was unfamiliar and frustrating to use, although there were some positive comments on presence and exciting interaction.

This analysis, in combination with the quantitative data for all groups, suggests that BB worked well for most groups, apart from a few poor performers who appear to have experienced more usability problems and become bored by the text-only interface. SL on the other hand worked well only when groups were motivated by the excitement and novelty of the 3D world, but for most, poor usability hindered its effectiveness.

Order effects
There was no evidence that the order in which the groups experienced the technologies or control condition affected performance; neither did the scenario order show any performance differences. However, the participants' comments did reveal interesting order effects. All six groups who encountered the FTF condition first commented that this allowed them to get to know each other quickly and help negotiate ways of collaborating. These collaborative processes enabled these groups to work more effectively in the technology conditions. Conversely, participants in seven of the groups who experienced one of the technology conditions commented that it would have helped to get to know each other first FTF. The following three quotes illustrate the advantages of meeting FTF before experiencing CMC technologies:

> We started from FTF; that made us more familiar with each other. It is better for us in the following discussions. That's why we seemed to be the first one to finish the tasks. The worst way is starting from SL as it is difficult to handle and it doesn't make sense to discuss in SL. I know there are differences between SL and BB but not that much different in these two tasks.

> Doing FTF first was very beneficial as we got to know and established a working process. Then we were able to take our online tasks in similar process to real life. I feel this allowed us to work quicker and more productively. So definitely it is beneficial in my opinion for a group to work together offline before online together.

> I had FTF, BB and SL because the FTF exercise was first; it was easier to organise a strategy, and it influenced the other exercise strategies and made it easier to carry them out. Because FTF was first, I understood the way the other participants thought and interacted through their body language and emotions, by interacting with them and understanding how to work with them in a group

These comments reflect the limitations of communication in the technologies which do not appear to have been mitigated either by avatars in SL or shared awareness functions in BB.

However there were a few comments which did favour the affordances of the technologies:

> Shyness and not knowing group members hindered. Easier to start discussion with complete strangers on BB and SL.

Learning and Reflection

Most of the groups (76%) reported that they did develop a common strategy to solve the problem and 4 out of 5 of the worse performing groups were among the groups without a common strategy. However nearly all the groups (86%) reported that their strategy did not change over iterations in the task, so there was little evidence of process learning. When asked in the final interview to reflect on how their performance could be improved, 19/21 of the groups suggested improvements. The majority concerned process (31%) concerned process such as better ways of structuring discussion, voting or agreeing prioritisation; 18% cited the need for more information on the problem domains, while 17% cited the technology and hindering performance, in particular the need for voice communication and better shared awareness of the problem space (item list and priorities agreed). Five groups reported social interaction could be improved (11% comments) while the remaining comments concern various issues such as the need for more time. Three of the bottom 5 performing groups encountered these social problems.

While learning may have emerged with more trials, the fact that most groups did form a strategy and the strategy did not change implies the technologies did not help learning or performance; furthermore, the reported social problems indicate that social awareness also needs to be improved.

In general reflections, technology was mentioned in 26% of comments spread across all groups. The need for better shared awareness and common ground was cited frequently; however, there are differing views on the affordances of technology compared to FTF, as illustrated in the following excerpts:

> Working with people who are complete strangers is much more convenient in SL and BB because the discussion will be more specific, focus, and aim to solve the problem. And not to be shy. BB is a pretty plain interface but more straight forward and easy to read. In SL users need to adjust location of the avatar properly in order to get a proper view, navigation error will caused the avatar to move out of sight. Some people are more comfortable to share information via online rather than FTF. In the chatting room we can chat and discuss anything.

I was more willing to argue my point in the FTF discussion whereas I was more inclined to avoid confusion in SL and BB hence exclude only the necessary opinions.

Overall neither technology provided effective affordances for collaboration and learning, and the groups did not appear to improve their modus operandi in spite of being aware that improvements could be made.

6 Discussion

Our findings demonstrate a positive outcome for the first research question: user experience with SL was richer than BB and more motivating; however, this did not produce better performance. So the answer to the second question we posed is that user experience in terms of affect and interaction does not appear to improve performance, although there was no difference between Blackboard and Second Life. The reasons for success or failure of the two technologies produced a complex picture. SL did not produce the expected benefit from the motivation of 3D interaction, probably because of usability problems encountered with the avatars. BB in contrast was perceived as being more usable, even though for some groups it was considered to be boring and not a stimulating user experience. FTF was expected to be most effective and indeed it was quickest and rated best on experience and positive emotions. However, FTF did not produce more accurate results, while BB did have a marginal, although non-significant, advantage. FTF may have been too easy for the participants so they did not pause to reflect and hence validate their solutions. In contrast, BB may have encouraged more reflection since communication was slower and messages were persistent.

Social relationships and overall group activities did not correlate with performance, so our results appear to agree with previous findings that social relationships have no positive effect on learning performance [15,24]. However, the degree of social closeness in most of our groups was limited, so our findings only relate to weak ties rather than stronger relationships.

The post-test attitude scores indicated a general pattern that BB was preferred for usability and overall, whereas SL was rated to be more exciting in terms of user experience. However, qualitative analysis of the five top and bottom performing groups showed that users may be divided into different cohorts: those who prefer the excitement of 3D experience and those who prefer more conventional task-oriented interfaces. The users' comments indicated that these reactions may have influenced performance in the tails of the distribution.

The avatars and affordances for collaboration provided by SL for the experimental task were limited. The avatars did not appear to motivate many users and did not provide a rich form of communication, contrary to the intuitions of media richness theory [4]. The numbered blocks did afford prioritisations but few users found this to be a natural form of collaboration. SL may be more advantageous in tasks where physical manipulation is necessary (e.g. construction, assembly and design). In the survival scenarios, the task was essentially abstract, and only involved discussion and

prioritisation. As some users commented, the avatars hindered communication, which evolved into a chat session. Use of voice in SL may have enhanced chat and possibly increased social presence; however, whether use of the voice modality would have enhanced the affordances of the avatars is an open question for future study. Theoretical models indicate that concurrent motor control (of avatars) and verbal communication can increase workload [32] so avatars in non-physical tasks may be more of a hindrance than a help.

BB was familiar for many of the participants, which may explain its advantage in usability and some of the adverse reactions to the less familiar SL. However, BB was compared unfavourably with FTF, because of its poor communication and weak affordances for collaboration. Topics threads in message exchanges were not used; instead, most groups used the chat facility and a Word document for shared awareness of their prioritised list. In conclusion, the affordances for collaboration provided by both technologies were limited, as demonstrated by the experimental order which enabled groups using FTF first to establish an effective modus operandi that they subsequently transferred to the technology conditions.

The significance of social awareness within groups was apparent from the importance attached to FTF interaction and the users' reflection on their experience. So while prior social relationships may not be necessary for collaboration [15,34], some social familiarity does appear to promote group interaction and performance. When technology is being used to mediate learning, a prior meeting FTF is advisable to help participants develop common groups so they can adapt their modus operandi to the technology. Although shared awareness was only cited by some of our users, it did appear implicitly in many reflections on the limitations of both SL and BB technology, reinforcing Neale et al.'s [16] view that such functions need to be improved.

The mixed methods approach we adopted showed the advantage of combining quantitative and qualitative data. The performance data and attitude scores set the scene for the overall capabilities of the technologies, but the reasons for success or failure needed qualitative data to discover the reasons for users' experience, which may have influenced performance. However, the experimental paradigm we used is still limited by the scope of tasks and experience that can be investigated, and longitudinal studies of collaboration in context are also essential for assessing the effectiveness of collaborative technologies [16]. Mixed methods approaches in experimental designs can however produce useful insight for comparing the affordances of collaborative technologies.

The implications of this study are, first, to caution against the rush to adapt apparently engaging technologies such as SL for collaborative learning, since we did not find the expected motivation bonus. There also appeared to be a usability penalty. However, these downsides might have been mitigated by more extensive training. The second implication concerns collaboration more generally, where the order effect, we found, indicates that giving groups FTF contact before interacting via CSCW/CSCL technology can help users develop a modus operandi for collaboration and adapt more effectively to the technologies. A third implication reflects limitations of the task. SL may be more acceptable and effective in tasks requiring physical interaction and hence active avatar roles, so it may be important to consider the collaborative task and learning objectives when introducing SL. In our future studies

we will explore different tasks and the effect of stronger social relationships where avatars might mediate a strong sense of social presence.

References

1. Baltes, B., et al. Computer-mediated communication and group decision making: A meta analysis, Organizational Behavior and Human Decision Processes, 87(1), 156–179 (2002)
2. Barron, B.: When smart groups fail, Journal of the Learning Sciences, 12(3), 307-359 (2003)
3. Brown, J., Collins, A., & Duguid, P.: Situated cognition and the culture of learning, Educational Researcher, 18(1), 32-42 (1989)
4. Daft, R.L., Lengel, R.H., & Trevino, L.K.: Message equivocality, media selection, and manager performance: Implications for information systems. MIS Quarterly, 11(3), 355-366 (1987)
5. Ducheneaut, N., Yee, N., Nickell, E., & Moore, R.J.: The life and death of online gaming communities: A look at guilds in World of Warcraft. Proceedings: CHI-07, pp. 839-848, ACM Press, New York (2007)
6. Dutson, A., et al.: Review of literature on teaching engineering design through project oriented capstone courses, Journal of Engineering Education, 86(1), 17-25 (1997)
7. Fjermestad, J.: An analysis of communication mode in group support systems research, Decision Support Systems, 37, 239–263 (2004)
8. Hertz-Lazarowitz, R., & Bar-Natan, I.: Writing development of Arab and Jewish students using cooperative learning (CL) and computer-mediated communication (CMC), Computers & Education, 39(1), 19–36 (2002)
9. Hew, K.H. & Cheung, W.S.: Use of three-dimensional (3-D) immersive virtual worlds in K-12 and higher education settings: A review of the research, British Journal of Educational Technology, 41(1), 33-55 (2010)
10. Hobman, E. et al.: The expression of conflict in computer-mediated and face-to-face groups, Small Group Research, 33, 439–465 (2002)
11. Irani, L. Hayes, G.R., & Dourish, P.: Situated practices of looking: Visual practice in an online world, Proceedings: CSCW 2008, pp. 187-197. ACM Press, New York (2008)
12. Janssen, J. et al.: Visualization of participation: Does it contribute to successful computer-supported collaborative learning, Computers & Education, 49, 1037–1065 (2007)
13. Johnsen, K., Raij, A., Stevens, A, Scott Lind, M.D. & Lok, K: The validity of a virtual uuman experience for interpersonal skills education, Proceedings CHI 2010, pp. 1049-1058. ACM Press, New York (2010)
14. Johnson, S., Aragon, S., Shaik, N., & Palma-Rivas, N.: Comparative analysis of learner satisfaction, and learning outcomes in online, and face-to-face learning environments, Journal of Interactive Learning Research, 11(1), 29-49 (2000)
15. Maldonado, H., Klemmer, S., & Pea, R.: When is collaborating with friends a good idea? Insights from design education, Proceedings: CSCL 2009. (2009)
16. Neale, D.C., Carroll, J.M., & Rosson, M.B.: Evaluating Computer-Supported Cooperative Work: Models and frameworks. Proceedings CSCW 2004 (pp. 112-121). (2004)
17. Neustaedter, C., & Fedorovskaya, E.: Capturing and sharing memories in a virtual world. Proceedings: CHI-09, pp. 1161 1170, ACM Press, New York (2009)
18. Newcomb, A., & Bagwell, C.: Childrens friendship relations: A meta-analytic review, Psychological Bulletin,117, 306–347 (1995)

19. Olson, G.M. & Olson, J.S.: Technology support for collaborative workgroups. In G.M. Olson, T.W. Malone, & J.B. Smith (Eds), Coordination theory and collaboration technology, pp. 559-584, Lawrence Erlbaum, Mahwah, N.J. (2001)
20. Ortony, A., Clore, G.L., & Collins, A.: The cognitive structure of emotions. Cambridge University Press, Cambridge (1988)
21. Pinelle, D., & Gutwin, C.: A review of groupware evaluations. Proceedings: IEEE WETICE, pp. 86-91 (2000)
22. Salmon, G.: The future for (second) life and learning, British Journal of Educational Technology, 40(3), 526-538, (2009)
23. Schroeder, R., Heldal, I., & Tromp, J.G.: The usability of collaborative virtual environments and methods for the analysis of interaction. Presence, Teleoperators and Virtual Environments, 15(6), 655-667 (2006)
24. Shah, P., & Jehn, K.: Do friends perform better than acquaintances? The interaction of friendship, conflict, and task, Group Decision and Negotiation, 2, 149-165 (1993)
25. Straus, S.: Technology, group process, and group outcomes: Testing the connections in computer-mediated and face-to-face groups, Human–Computer Interaction, 12, 227-266 (1997)
26. Strauss, A., & Corbin, J.: Basics of qualitative research, Sage Publications, London (1998)
27. Sutcliffe, A.G. Designing for user engagement: aesthetic and attractive user interfaces. In J.M. Carroll (Ed.), Synthesis lectures on human centered informatics. San Rafael CA: Morgan Claypool (2009)
28. Thompson, L., & Coovert, M.: Teamwork online: The effects of computer conferencing on perceived confusion, satisfaction and postdiscussion accuracy, Group Dynamics, 7, 135-151 (2003)
29. Tractinsky, N., Katz, A.S., & Ikar, D.: What is beautiful is usable. Interacting with Computers, 13(2), 127-145 (2000)
30. Tractinsky, N. & Zmiri, D. Exploring attributes of skins as potential antecedents of emotion in HCI. In P. Fishwick (Ed.) Aesthetic computing. Harvard: MIT Press (2005)
31. Tromp, J.G., Steed, A., & Wilson, R.: Systematic usability evaluation and design issues for collaborative virtual environments, Presence: Teleoperators and Virtual Environments, 12(3), 241-267 (2003)
32. Wickens, C. Multiple Resources and Performance Prediction. Theoretical Issues in Ergonomics Science 3, 2, 150-77 (2002)
33. Witmer, B.G., & Singer, M.J.: Measuring presence in virtual environments: A presence questionnaire. Presence, 7, 225-240 (1999)
34. Zajac, R., & Hartup, W.: Friends as co-workers: Research review and classroom implications, The Elementary School Journal, 98, 3-13 (1997)