SimCoL – A Simulation Tool for Computer Supported Collaborative Learning

Nobel Khandaker and Leen-Kiat Soh

***Abstract*— Researchers designing the multiagent tools and techniques for CSCL environments are often faced with high cost, time, and effort required to investigate the effectiveness of their tools and techniques in large-scale and longitudinal studies in a real-world environment containing human users. Here, we propose SimCoL, a multiagent environment that simulates collaborative learning among students and agents providing support to the teacher and the students. Our goal with SimCoL is to provide a comprehensive testbed for multiagent researchers to investigate (1) theoretical multiagent research issues e.g., coalition formation, multiagent learning, and communication, where humans are involved, and (2) the impact and effectiveness of the design and implementation of various multiagent-based tools and techniques (e.g., multiagent-based human coalition formation) in a real-world, distributed environment containing human users. Our results show that SimCoL (1) closely captures the individual and collective learning behaviors of the students in a CSCL environment, (2) identify the impact of various key elements of the CSCL environment (e.g., student attributes, group formation algorithm) on the collaborative learning of students, (3) compare and contrast the impact of agent-based vs. non-agent-based group formation algorithms, and (4) provide insights into the effectiveness of agent-based instructor support for the students in a CSCL environment.**

***Index Terms*—Collaborative Work, Cooperative Systems, Educational Technology, Simulation.**

I. Introduction

COMPUTER-SUPPORTED collaborative learning (CSCL) environments facilitate student learning by enhancing their collaborative learning using computer and Internet technologies. Today, CSCL environments contain agents and agent-based services to improve the collaborative learning of students from two different aspects. First, the agents act as assistants to the students by monitoring the difficulties they face and helping them with customized support. Second, the agents act as assistants to the teacher providing decision support and helping him or her with tasks like group formation. To design agents, agent-based services, and agent-based algorithms for a CSCL environment, it is essential to: (1) understand how those various elements of the CSCL environment work together to produce the learning outcome of the students and (2) investigate how those services impact the students' interactions and learning outcomes. Furthermore, without testing their algorithm on a large group of students for a sufficiently long time, it is difficult for the researchers to: (1) fully understand the impact of their designs and (2) evaluate their designs and algorithms against the state of the art. Albeit considered the most authentic way of validating the results, it is often difficult to conduct experiments with human users for various reasons: (1) it is difficult to acquire enough students for long enough time to do the experiments, (2) replication of experiments is often not possible, and (3) experiments may yield unwanted consequences (e.g., student apathy toward the use of CSCL environment) if the agents or agent-based services do not work as expected. These issues can be alleviated by using *agent-based simulation*.

However, the existing tools designed to simulate the CSCL environment has yet to consider the role of agents in supporting (or scaffolding) the activities. When designed based on the individual and collaborative learning theories, the students and their interactions with each other in the simulation would closely represent the collaborative learning in the real-world CSCL classroom. Existing tools such as [5] only simulate the student behavior using agents and do *not* include agents that act as the assistant agents or any agent-based services or algorithms. As a result, the decision making process of the CSCL module that provides scaffolding to help both the teacher and the students, as well as the appropriateness and costs of such a module, have not been studied comprehensively.

In this paper, we describe SimCoL—a *multiagent application for simulating the collaborative learning of a set of students in the CSCL environment*. The *inspiration* source of our paper is CSCL environments that combine research ideas from psychology (especially educational psychology), education, and computer science to create an online collaborative learning environment for students. This simulator would *improve* the computer-supported collaborative learning, and collaborative learning in general, in the following ways. First, SimCoL would allow researchers and teachers to gain insights into the collaborative learning process by carrying out *what-if* simulations that reveals the *emergent* outcome for a given environment setting (e.g., students with specified knowledge and ability). Second, SimCoL would allow the researchers and teachers to better understand the impacts of the administrative decisions like (1) group formation scheme, (2) group size, (3) and agent-based support [28] on the student learning outcome.

To show the validity of SimCoL and to illustrate the various

Manuscript received August 30, 2009. This work was supported by NSF Grant DMI-044129 and Microsoft Research.

Nobel Khandaker is a doctoral candidate in the Dept. of Computer Science and Engineering in the University of Nebraska – Lincoln, NE 68588 USA.

Leen-Kiat Soh is an associate professor in the Dept. of Computer Science and Engineering in the University of Nebraska – Lincoln, NE 68588 USA.

scenarios that SimCoL could be used to investigate the sensitivity of the CSCL environment design impacting various aspects of student models, we have run several large simulations. Our results show that the SimCoL environment is able to capture the change in the knowledge gain of the students due to: (1) the changes in the attributes (e.g., ability and motivation) of the participating students and (2) the various techniques (e.g., group formation method) used in CSCL. Further, the individual and collective learning behavior patterns of the students in SimCoL *closely* represent the learning behavior patterns reported by the CSCL researchers. These results suggest the usefulness of SimCoL as a simulation tool.

The rest of the paper is organized as follows. Section II presents a set of learning theories and observations based on the individual and collaborative human learning process and define the scaffolding of students in a collaborative learning environment. In Section III, we use the observations in Section II to design the agent that represents the teacher, the agents that represent the students, and the tasks in the SimCoL environment. Then we describe how the SimCoL environment was realized using Repast—a multiagent simulation tool, in Section IV. Section V describes our experiment setup and results. Finally, in Section VI, we present some related work and in Section VII discuss the conclusions.

Note that we use the following terminologies in this paper. An agent that simulates the teacher's behavior in SimCoL is a *simulated teacher*, while one that simulates a student's behavior a *simulated student*. An agent that assists the students in forming groups is a *student-assistant agent*, while one that assists the teacher a *teacher-assistant agent*.

II. Learning

In this section, we discuss definitions, theories, and empirical observations regarding three different aspects of students' learning processes: (1) individual learning, (2) peer-based learning, and (3) collaborative learning in student groups. Using these learning theories, we derive a set of observations that are used in Section III to build agents to simulate the student collaborative learning behavior in a CSCL classroom.

A. Individual Learning

We use "learning" to refer to the improvement in a student's knowledge or expertise on a topic or skill, which could be topic-specific, e.g., learning how to solve differential equation, or topic-independent, e.g., teamwork or communication skills. According to learning theories [1], [7] the four main elements that affect how a person learns are: (1) what the student already knows (knowledge), (2) how able/intelligent the student is (ability), (3) how motivated the student is (motivation), and (4) the emotional state of that student (emotion). The cognitive components that represent these factors are: (1) the crystallized intelligence as accumulated knowledge stored in long-term memory, (2) fluid intelligence as represented by working memory capacity, and (3) motivation as represented by working memory allocation [1], and (4) emotional state [7]. Next, we define these elements in greater detail.

Shell and Brooks [1] use the term knowledge to refer to the accumulated knowledge in a student's long-term memory. The ultimate result of learning would occur as the improvement of the knowledge of the students. Shell and Brooks [1] use ability to represent the cognitive ability or intelligence of a person. They suggest that there are two different parts of ability: fluid intelligence and crystallized intelligence. The fluid intelligence is a fixed entity that deals with general cognitive capacity and crystallized intelligence represents the accumulated knowledge of the student. Furthermore, the fluid intelligence is basically the working memory of a student [1]. However, there is a difference between the absolute working memory capacity a person has and the amount of working memory capacity he or she has available at a particular time for a particular task. The behavior of a person while working on a task and the improvement in his or her knowledge due to learning by working on that task depend upon the amount of working memory that person has available at that time. Further, the amount of working memory available to any person at a time is determined by: (1) his or her existing knowledge for that task, (2) his or her motivation to work on that task, and (3) emotion [1]. Motivation determines why we do what we do [1]. In other words, motivation is the process whereby goal directed behavior is instigated and sustained. Finally, the emotion of a student determines whether the students are feeling happy or sad. So, we write our first observations as:

Observation 1: *A student's improvement of knowledge of a topic is mainly affected by: (1) his or her existing knowledge, (2) ability, (3) motivation, and (4) emotion.*

Observation 2: *The amount of working memory available to a student determines how much he or she can learn*.

Observation 3: *The working memory of a student interacts with his or her prior knowledge and new information (regarding a task) to produce learning and behavior*.

Observation 4: *A student's available working memory for a task can be described as his or her ability for that task*.

Although the aforementioned four components that affect learning are cognitively distinct from one another, there are combinatorial effects [1]: (1) the prior knowledge stored in the long-term memory interacts with the working memory to produce learning, (2) available amount of working memory limits how much prior knowledge and information can be used/activated at any time, (3) the amount of working memory is determined by motivation, extent of prior knowledge, and emotion, and (4) as knowledge increases, it increases the effective working memory capacity allowing acceleration of future learning processes. Finally, according to the recent research work on perceptual and motor acquisitions [26]-[27] the pace of skill acquisition for a learner accelerates in the beginning and slows down to a stable state, leading to:

Observation 5: *A student's available working memory for a topic is proportional to his or her: (1) knowledge on that topic, and (2) motivation to learn that topic. Furthermore, this available working memory is inversely proportional to the emotional state of that student*.

Observation 6: *As the knowledge of a student on a particular topic increases, his or her learning outcome for that topic* *would increase at the beginning and slow down to a steady state after a certain amount of time.*

B. Peer-Based Learning

When a student is working with his or her peer to solve some assigned task, the student and the peer may learn from each other about that task. The possible learning scenarios between two interacting peers are summarized by [8] such as (Table I): learning by *observation*, learning by *teaching/guiding*, learning by *being taught*, learning by *reflection/self-expression*, learning by *apprenticeship*, learning by *practice*, and learning by *discussion*. From these peer-based learning scenarios, we observe that the prior knowledge of the participating students plays an important role in deciding what type of learning scenarios may occur. For example, learning by teaching (and learning by being taught) is more common among two students where one student with more prior knowledge teaches his or her peer who has less prior knowledge. Furthermore, the *difference* between two interacting students' prior knowledge about how to solve a certain task can hinder their learning. This effect is described in Vygotsky's zone of proximal development (ZPD) theory [9]. For example, it may be difficult for two students to learn from each other if the amount of prior knowledge they have on a topic is vastly different from each other [9]. So we write:

Observation 10: *Two students may learn about a topic from their interactions (Table 1) when the content of prior knowledge they have are not too different from one another.*

TABLE I
POSSIBLE LEARNING SCENARIOS AMONG PEERS

Observation	**Student—Peer Know.**	**Learning by**
7	High—High	Observation, Reflection, Practice, & Discussion
8	High—Low or Low—High	Observation, Teaching, Being Taught, Reflection, Practice, & Discussion
9	Low—Low	Observation

C. Collaborative Learning

The term "collaborative learning" is an instruction method in which students at various performance levels work together in small groups toward a common goal [10]. Derived from Stahl [11] are:

Observation 11: *The collaborative knowledge building is a cyclic process that feeds on itself.*

Observation 12: *This collaborative knowledge building cycle is a hermeneutic cycle, meaning, "one can only interpret what one already has an interpretation of".*

Observation 13: *Individual knowledge of a student is gained from collaborative knowledge of his or her group members through interaction. That collaborative knowledge is in turn produced by individual knowledge of the interacting group members.*

Kreijns [12] describe the interaction between students as the key to collaboration among group members. Furthermore, researchers [25] suggest that collaborative learning occurs from the exchange of dialogues among the students.

Observation 14: *The collaboration among the members of a group of students occurs due to their interaction/discourse with each other.*

Zumbach [13] describes a collection of dyadic (between two students) interactions for a group of students which were reported by researchers in the CSCL community. An example of interactions mentioned in [13] is: (a) student a proposes a solution for the assigned task, (b) student b accepts or proposes another solution to the task. Thus:

Observation 15: *The compilation of discourse/interaction patterns presented by Zumbach et al. [13] describes a typical dyadic (between two students) learning scenario in terms of a chain of action-reaction patterns.*

The quality the discourse/interactions within a group depends on the affective state of a student [2] and his or her social relationship with other students in the group. Jones and Issroff [14] and Vass [15] report that, students who are friends have established ways of working which are implicitly understood rather than explicitly discussed. In addition, [12] mentions that social relationships contribute to common understanding, an orientation towards cooperation, and the desire to remain as a group. Finally, as reported in [3], the students form their view of other students due to the type and extent of collaboration they receive from their peers. Clear and Kassabova [16] further report that in collaborative learning settings it is common to have students whose motivation is affected by the motivation of other group members. When the other group members are motivated to learn and to collaborate, it increases the motivation of a student who had low motivation when he or she joined the group, and vice versa. We derive from the above the following observations:

Observation 16: *Good social relationship improves the quantity and quality of interactions among group members.*

Observation 17: *The quantity and quality (i.e., learning outcome) of interactions among a group of students vary over time due to factors internal and external to the classroom environment. Improvement in social relationship among the members of a group improves the quality of collaborations among them. On the other hand, when a student group member experiences distracting factors, that experience reduces the quality of his or her collaboration with other members.*

Observation 18: *Motivation of the group members' impacts the motivation of a student positively and negatively.*

Observation 19: *Social relationship between a student and his or her peer (as perceived by the student) change according to the frequency, extent, and quality of collaboration (e.g., how many times did my peer helped me).*

D. Scaffolding

Bruner [17] and Cazden [18] define scaffolding as the act of providing assistance to a child so that he or she is able to carry out a task (e.g., solve a problem) that he or she cannot do by herself. Over time, the concept of scaffolding has been introduced into traditional classrooms to aid learners to achieve difficult learning objectives and complete difficult tasks [3] where tools and software are used to (1) offer structure and support for completing a task and (2) promote peer interactions to enable peers to support each other's learning. In the first type of scaffolding, the students are provided information about how to better approach to solve the task that they are having difficulty with. In the second type of scaffolding, the peer support of a student is enhanced in the hope that those peers would provide guidance and information for that student to help him or her solve that task. Researchers in the CSCL community are now utilizing scaffolding in the form of incorporating structure of learning activities (e.g.,[19]) and improving peer support (e.g.[20]). As CSCL researchers (e.g., [3, 20] note that due to being in different zones of proximal development, the learners benefit most when the scaffolding is targeted toward their zone of development. So, one of the recommendations provided to the CSCL practitioners is to customize the scaffolding to specific learners' needs. Hence:

Observation 20: *Scaffolding in the CSCL environment can be provided by: (1) providing structure and support for completing tasks and (2) improving of peer support.*

Observation 21: *Scaffolding in the CSCL environment may be used to improve the knowledge of the learners regarding the assigned task.*

Observation 22: *Learners in a CSCL environment benefit more when the provided scaffolding is targeted to their zone of proximal development.*

III. SimCoL Environment

The SimCoL environment represents a CSCL environment where the teacher forms student groups and assigns a set of tasks and the students solve those tasks collaboratively to improve their knowledge about some topic. The SimCoL environment is defined as a 3-tuple: . Where is a set of tasks, is an agent who simulates the teacher, and is a set of agents who simulates the students in a collaborative classroom environment. In this section, we first define the tasks . Then, based on the observations presented in Section II, we describe the attributes and the behavior of agents who represent the students in SimCoL. Furthermore, we describe how the simulated teacher forms groups of simulated students and carries out CSCL classroom sessions in the SimCoL environment using a set of simulation steps. Finally, we describe the collaboration process of the simulated students in a group in SimCoL using a set of simulation steps and discuss how their attributes change.

A. Task

The tasks in SimCoL represent the problems and exercises that are solved by the students in a CSCL environment. The set of tasks is denoted by, where,

(1).

Here, denotes the concept of the task. This concept represents the subjective knowledge required to solve the task. , is the difficulty of the task as determined by the simulated teacher. is the time limit within which the task is to be completed. where is a vector representing the simulated student groups' (who are working on the task) view of the solution quality of the task at time .

B. Simulated Student

We represent the model of each simulated student in SimCoL by a 6-tuple:

(2),

where, is the knowledge of simulated student at time with representing the concept of and is the expertise, i.e., the amount of knowledge the simulated student has about the concept. The goal of simulated student collaboration is to increase the value of this expertise. , is the ability of at time for task . , is the motivation of at time . , is the emotional state of simulated student at time where is the social relationship between and at time as perceived by . denotes the target solution quality of of at time .

We have included , , and in the model according to Observation 1 and included according to Observations 16 and 17. Also, combining Observations 4 and 5, we assume that the ability of a simulated student is related to his or her knowledge, motivation, and emotion in the following way:

(3), where and are weights. According to Eq. 3, the ability of a simulated student for a particular task at any time is proportional to the sum of his or her expertise on the concept of that task and motivation minus the absolute value of his or her emotional state. We also define the target solution quality of a simulated student with: (4). So, a simulated student's target of the quality of the solution of the assigned task is proportional to his or her ability for that task. According to Observations 3 and 4, the ability of a simulated student determines how much of his or her existing knowledge can be activated to produce behavior (i.e., effort to solve the task) and learning. Therefore, given the same time limit for a task , a simulated student with higher ability would be able to solve the assigned task better than a simulated student with lower ability. So, we assume that the simulated students have targets of the final solution quality according to their own abilities.

C. Simulated Teacher

The teacher in SimCoL acts as the coordinator of the CSCL sessions. The teacher delivers instructions, forms groups, and assigns collaborative tasks. In SimCoL, we have implemented three different group formation methods: **random**, **Hete-A** [21], and **VALCAM** [22] group formation method. Table 2 shows how the teacher carries out the CSCL session through a set of simulation steps. First, the teacher initializes the classroom (tasks, group formation scheme, how often scaffolding should be provided, and how many groups would receive scaffolding). Then, for each initialized task, the teacher: (1) initializes a collaborative session (Step 2a), forms simulated student groups (Step 2b-d), and announces the start of the collaborative session to all simulated students (Step 2e). Then until the collaborative session is over, the teacher periodically sorts the groups according to their current achieved solution quality of the task (Step 2fa(1)) and then selects the groups who have the lowest solution quality. Those selected groups are then provided scaffolding (Step 2fa(2)). Finally, the teacher announces the end of the collaborative session when the time limit for the current task is over (Step 2g).

TABLE II
SIMULATION STEPS OF TEACHER

1. Initialization: , group formation scheme, scaffolding period, , simulated students , and agents
2. For all tasks , do,
   a. Initialize collaborative Session : , , , and Announce task to ,
   b. If , form Random Group for
   c. Else If , form Hete-A [21] groups for
   d. Else If , form VALCAM [22] groups for
   e. Announce start of collaborative Session to
   f. While (true)
      a. If
         1. Sort (ASC) according to
         2. For
            Provide scaffolding to
      b.
   g. Else Announce end of collaborative Session to

D. Assistant Agents

The student-assistant and teacher-assistant agents have been incorporated in SimCoL to implement various agent-based coalition formation algorithms. Each student-assistant agent in SimCoL is assigned to a simulated student and it monitors the change in that assigned simulated student's: (1) expertise gain and (2) social relationship with other students. The teacher-assistant agent is assigned to the instructor to (1) assign and monitor student collaborative performances and assign them virtual currency according to that performance and (2) communicate with the student agents to form groups using VALCAM [22].

E. Collaboration and Scaffolding

Following Observations 14 and 15, in SimCoL, we simulate the collaborative behavior (i.e., collaboration to solve the assigned task and to improve expertise) of a group of simulated students using a series of *dyadic* interactions among the group members. Here, we describe how those interactions occur in SimCoL. First, we define the following functions that dictate the behavior of the student agents simulating the collaborative learning in SimCoL. Here, we assume that two simulated students and with models and are working in a group to solve task and all variables are weights:

Motivation Update (Observation 18):

(5)

Collaboration Probability (Observation 16):

(6)

Collaboration Cycle (Observation 15): denotes a collaboration cycle completed by with at time for task . Here, denotes an utterance of action, denotes an utterance of reaction in reply to the action , and denotes the reaction in reply to the reaction . denotes a collaboration cycle initiated by but declined by . is the set of all collaboration cycles between and for .

Solution Quality Update: If and *Otherwise* (7), where , denotes the solution quality update probability threshold and a random number that is drawn from a uniform random distribution respectively.

Human Expertise Update (based on Obs. 3,4,7-10 & 13): if otherwise

(8) with (9) where is the ZPD constant.

Social Relationship Update (based on Observation 19):

(10)

Scaffolding Effect (based on Obs. 20-22): If and *0* otherwise (11) where is the scaffolding object, , denotes the level of expertise for the simulated student the scaffolding is designed for, denotes the cost (e.g., time and effort required to design the object) of the scaffolding, is a probability value drawn from a uniform distribution, and is scaffolding threshold.

Table III shows the simulation steps of a simulated student in SimCoL with the various formulas that are used by the agents in parenthesis. During initialization, the simulated student receives its group assignment and the task (Step 1) from the simulated teacher (Step 2a in Table II). Then the simulated student updates its own motivation according to other group member's motivations, and its ability. During the session, the simulated student tries to collaborate with its group members if the quality of the solution is less than its expected solution quality (Step 2a) or if someone else in the group wants to collaborate (Step 2b). In both of these cases, whether the collaboration is successful or not depends on the collaboration probability (Step 2b(i)). During the collaborative session, if the simulated student receives scaffolding from the simulated teacher (Step 2c) in the form of a scaffolding object, it updates its expertise. When the collaborative session ends, the simulated student updates its own view of social relationship with its group members (Step 3).

TABLE III
SIMULATION STEPS OF STUDENT

Simulation Steps of Student
1. Initialize: group , task , update motivation (5) and ability (3)
2. Until collaborative Session is over, do,
a. If Then
i. Propose collaboration to randomly chosen
ii. If agrees
Complete and store , update solution Eq. 7,8
iii. Else
Store failed collaboration cycle in

b. If received collaboration request from Then
i. If Then
Complete and store , and update expertise (Eq. 8)
ii. Else
Decline request from and store failed cycle in
c. If received scaffolding , then
Update expertise (11)
3. Update social relationship (10) for group members

IV. Implementation

The SimCoL environment was implemented using the Java version of the Repast [24] – a multiagent simulation toolkit. Table IV describes: (a) the categorizations and the ranges of the randomly generated values in SimCoL, i.e., the student attributes and the weights and constants used in the equations in Section III. Fig. 1 shows the deployment diagram and Fig. 2 shows the input/output/control parameters of SimCoL.

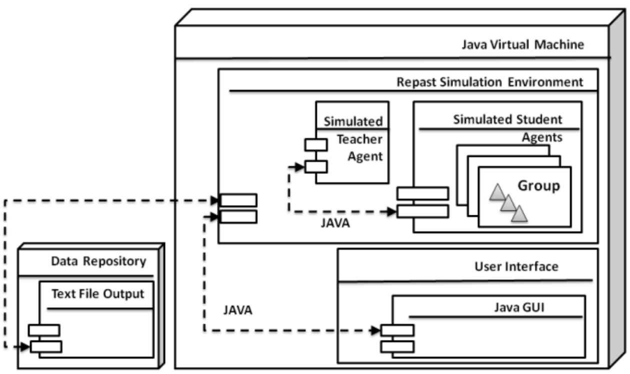

Fig. 1. Deployment diagram of SimCoL.

TABLE IV
CATEGORIZATIONS, DISTRIBUTIONS, WEIGHTS, AND CONSTANTS

Eq.	Attribute	Categorization	Generated from Normal Distribution with
1	Task Difficulty	Low , erate , and high	, , and range
2	Expertise	Low , moderate , and high ,	, , and range
2	Ability	Low , moderate , and high	Calculated using 0 with range
2	Motivation	Low , moderate[0.2,0.8), and high	, , and range
2	Emotion	Sad , tral , and py .	and , , and range
2	Social Relationship	Unknown , familiar , and friend	, , and range
	Weights and Proportionality Constants		
3	Weights: , , and		
4	Prop. constant:		
5	Weights: and		
6	Weights: and		
7	Prop. constant:		
8	Weights: and , prop. constant:		
9,10	Proportionality constant:		
	Other Constants		
Collaboration threshold			
Zone of proximal development threshold			

V. Results

The goal of our experiment is three-fold: (a) discussing how SimCoL is able to identify and reveal the complex relationship between the variables (i.e., student attributes) of a computer-supported collaborative learning environment, (b) comparing the emergent phenomenon of student performance in SimCoL with that of the published CSCL results, and (c) providing evidence of the validity of SimCoL simulation environment. In subsection A, we discuss how the social relationships among the students in SimCoL impact their collaborations. In subsection B, we describe the experiment that shows the interdependence of the students' attributes on their collaborative learning outcome. In subsection C, we study the impact of group formation and group size on student learning.

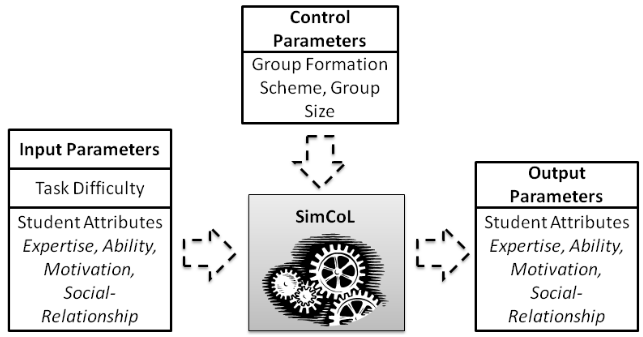

Fig. 2. Input Output and Control Parameters of SimCoL.

This allows us to: understand the usefulness of SimCoL in carrying out what-if scenarios in CSCL environments and correlate the observed patterns of student behavior in SimCoL with that of the reported CSCL studies. In subsection C, we compare and validate the emergent patterns of student behavior in SimCoL with that of the observed student behaviors in the reported CSCL studies. Notice that, all of our experiments are replicated for 10 simulation seeds.

A. *Learners' Collaboration Work*

Here, we ran the simulation for students for simulation ticks for each run by varying the values of two attributes at a time. We then plotted the successful collaborations of the students against their changing attribute values. Among *all attributes*, we have found that the social relationship among the group impacts students' collaboration efforts the most. Fig. 3 shows the results and Table V shows the skewness and kurtosis values.

This indicates that as the collaborative learning researchers [16] mention, social relationship among the students is a critical factor in improving the collaborations among them. Furthermore, the lack of the strong relationship between the other attributes like expertise can be explained by our formulation of collaboration probability (Eq. 6). The two key factors that determine a student's participation in a collaboration cycle is the target solution quality (Eq. 4) and social relationship. However, if the task solution quality is high (due to other members' contributions), a student's expected solution quality is then mainly determined by his or her social relationship with other group members. This result portrays a common scenario where students often refuse to collaborate/contribute when they see other members solving the task [29].

B. Compound Impact Analysis

This compound impact analysis allows us to: (a) investigate how the students belonging to the different categories of an attribute respond to the changes in another attribute, e.g., how do the student with low expertise react to a change in their motivation, and (b) investigate whether a student's lower value in an attribute can be compensated by a higher value. For this experiment, we ran the simulation for students for simulation ticks for each run by varying the values of two attributes at a time. Fig. 4 and Fig. 5 show the average and standard deviation of student expertise gain for students with low, medium, and high expertise against changing motivation. Fig. 6 and Fig. 7 show the average and standard deviation of student expertise gain for students with low, medium, and high expertise against changing motivation. Table VI and VII show the skewness and kurtosis of the students with low, medium, and high expertise plotted in Fig. 4 and Fig. 6.

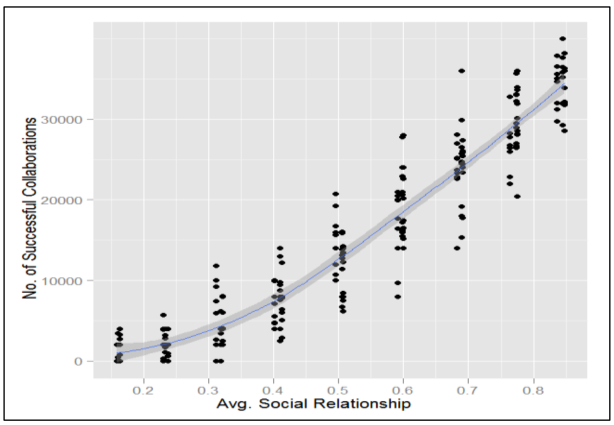

Fig. 3. Successful student collaborative cycles vs. average student social relationship.

TABLE V
SKEWNESS AND KURTOSIS OF DISTRIBUTIONS OF COLLABORATION CYCLES

Social Relationship	0.2	0.4	0.6	0.8	1.0
Skewness	1.2	1.8	0.2	-0.1	-0.8
Kurtosis	3.2	6.4	2.1	2.3	3.3

According to Fig 4, we see when the average motivation of the students is increased, the students of all categories (low, medium, and high) of expertise are able to improve their expertise gain and there are students who fall behind (unchanged standard deviation). This is to be expected as dictated by expertise update equation Eq. 8 where the expertise increase is determined by the motivation and difference in expertise. Furthermore, the unchanged standard deviation indicates that there are students in all three cases (low to high motivation) who cannot gain expertise due to the increased motivation.

Fig. 6 shows that as the social relationship of students improve, their expertise gain improves at first, and then that rate of improvement slows down to zero. Furthermore, Fig. 7 shows that the standard deviation of the students expertise gain remains somewhat unchanged with the increasing social relationship. This occurs due to our use of student social relationship while calculating the collaboration probability among two students (Eq. 6). The expertise gain of the students in the group depends on how well they collaborate. As the social relationship among the students starts to increase from initial lower value, the probability of them collaborating increases. As a result, they are able to gain more expertise. However, when their social relationship values are near maximum and all students in every group are collaborating, increase in the social relationship further, does not impact their expertise.

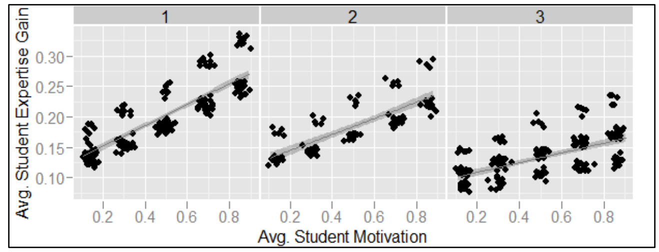

Fig. 4. Avg. student expertise gain vs. average student motivation for low, medium, and high expertise (left to right) students.

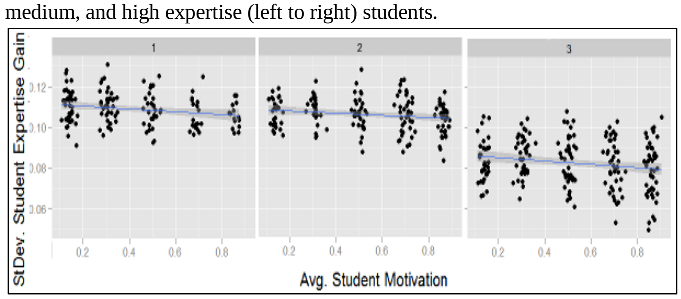

Fig. 5. StDev. of student expertise gain vs. average student motivation for low, medium, and high expertise (left to right) students.

TABLE VI
SKEWNESS AND KURTOSIS OF DISTRIBUTIONS OF EXPERTISE GAIN (FIG. 4)

	Low	Medium	High
Skewness	0.6	0.5	1.3
Kurtosis	2.3	3.0	4.9

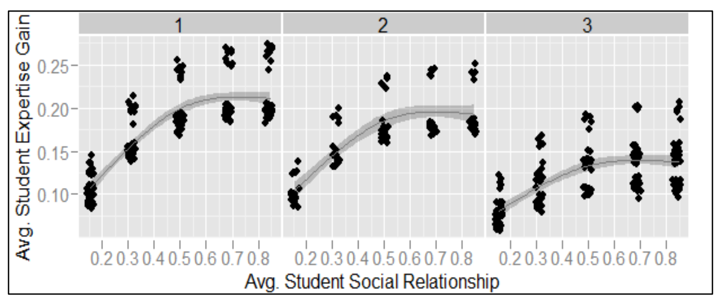

Fig. 6. Average student expertise gain vs. average student social relationship for low, medium and high expertise (left to right) students

TABLE VII
SKEWNESS AND KURTOSIS OF DISTRIBUTIONS OF EXPERTISE GAIN (FIG. 6)

	Low	Medium	High
Skewness	0.13	0.1	0.9
Kurtosis	2.1	3.2	3.5

Our observations here provide us the insight that, the critical student attributes in a CSCL setting often impact (negatively and positively) one another's contributions to a student's collaboration and learning. This observation is in sync with the current theories that describe the collaborative learning mechanism being affected by a variety of student attributes like motivation [1,2,16]. Thus, while setting up the collaborative learning environment, or when evaluating the outcome, it is important to look at all of those critical attributes together instead of in isolation as often discussed in the results of current CSCL research [13, 14, 21, 22]. In other words, while determining the impact of a collaboration script, group formation scheme, or other CSCL tool, the students' learning outcome alone may not be a sufficient indicator. Instead, we should also look at factors like motivation and social relationship that could have *influenced* the students' expertise gain.

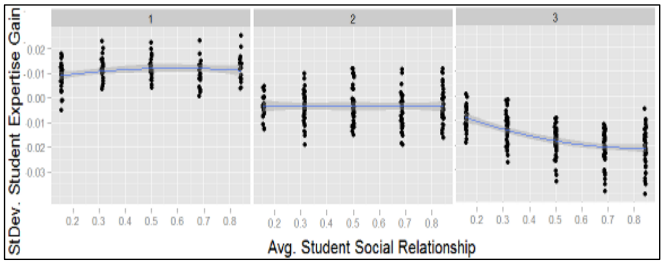

Fig. 7. Standard deviation of student expertise gain vs. average student social relationship for low, medium and high expertise (left to right) students.

C. Impact of Group Formation Method

In this section, we study the effect of two teacher-controlled aspects of a typical CSCL environment, i.e., (1) the group size and (2) the group formation scheme, on the average expertise gained by the students. During the simulation, the student groups in this experiment were formed using **Random**, **Hete-A**, and **VALCAM** group formation methods with the group size selected from the range of]. VALCAM is an agent-based algorithm of group formation which uses a multiagent system to form student groups that brings together experts with non-expert students where the members have high social relationships. Hete-A algorithm is a non-agent-based algorithm that forms heterogeneous groups. In Hete-A, the students are first categorized by assigning them to a matrix whose dimensions represent the attributes of a student. Once the students are categorized, the Hete-A algorithm builds heterogeneous groups by selecting students with the highest difference of attribute values according to their position in the matrix. Here, the Hete-A algorithm was used with the motivation and expertise as the two matrix dimensions. We first ran the simulation with the parameters described in Table IV for 30 students for 2000 ticks with expertise distribution mean , expertise distribution standard deviation , and collaboration threshold , for a set of 30 students, for different tasks and for simulation ticks, where the students mean expertise and social relationship was set to the mean initial values reported in [22]. Fig. 8 shows that the students in the VALCAM-formed groups performed better than the randomly formed and HETE-A formed groups.

The improvement in student performance in VALCAM-formed groups was reported in [22], so this result reproduces those observations. This improvement of student performance in VALCAM-formed groups can be explained by the way VALCAM forms student groups that contain expert and non-experts who have high social relationships amongst themselves. Since, the collaboration probability (Eq. 6) and therefore the collaborative learning in SimCoL is determined by the expertise difference (Eq. 8) and social relationship (Eq. 6), VALCAM-formed groups in SimCoL were able to collaborate better (i.e., *higher* number of successful collaborative cycles) yielding higher collaborative learning outcome. These results suggest that by setting the initial classroom conditions (e.g., student attributes) in SimCoL like a CSCL classroom, we could execute *what-if* scenarios by running simulations and compare the performances of group formation mechanisms.

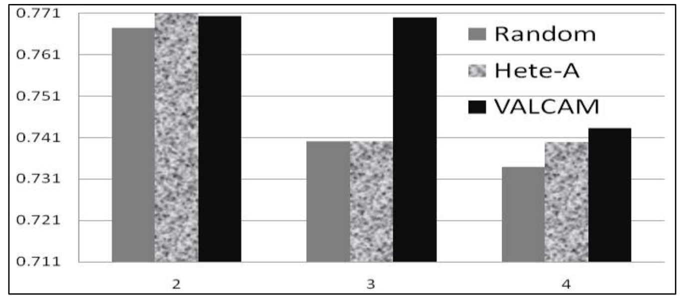

Fig. 8. Average expertise gain (y-axis) for varying group sizes (x-axis).

D. Cost and Expertise Gain through Scaffolding

In this experiment, we investigate how the individual and group scaffolding improves the expertise of the students when they are collaborating in various types of groups. To collect data, we ran the simulation with the same default set of parameters Table IV for students for simulation ticks. We calculated: (1) the average improvement in the expertise gain of the students and (2) the cost incurred for providing scaffolding for individuals and groups. For a group in this experiment, one scaffolding object is used *per group* for group scaffolding (i.e., scaffolding cost is required for one scaffolding object) and one scaffolding object *per group member* (i.e., scaffolding cost is equal to the sum of all generated scaffolding objects) is used for individual scaffolding. Fig. 9(a) shows the average improvement of student expertise gains of the students when they are working in random, Hete-A, and VALCAM formed groups. Fig. 9(a) shows that the students in all groups are able to improve their expertise more from the individual scaffolding than from the group scaffolding. This is expected, since: (1) individual scaffolding addresses individual students' needs, and (2) according to our design of scaffolding (Eq. 11), a student's expertise is improved most when the scaffolding is targeted towards his or her expertise level.

Fig. 9 (b) shows that for all three types of groups, the group scaffolding yielded more expertise gain per unit cost than the individual scaffolding. The cost of scaffolding denotes the time and effort required for providing scaffolding to the students. Providing individual scaffolding requires more cost since each individual student has to be modeled and different types of scaffolding have to be provided to the students according to their expertise level. On the other hand, group scaffolding requires less cost since the scaffolding action is more generic and only one type of scaffolding is provided to the entire group. But unexpectedly, *the group scaffolding is shown to be more economical in terms of expertise improvement per unit cost*. Upon closer analysis, this can be explained by the *cyclic* and *convergent* nature of the collaborative knowledge building process (Observation 11). Due to this cyclic nature, collaborative knowledge is transferred among the group members due to their interactions throughout the collaborative session. Furthermore, our non-adaptive scaffolding process periodically provides scaffolding to a fixed number of student groups by first sorting them according to their performances. However, near the end of the collaborative cycle, due to the heterogeneous nature of groups of the random, Hete-A, and VALCAM groups, there are some students who have already reached near-maximum expertise level. So, scaffolding for such group members is no longer effective. As a result, both individual and group scaffolding do not yield any expertise improvement for those high-expertise group members. But, for those high-expertise group members, the individual scaffolding incurs a much higher cost than would the group scaffolding. As a result, the improvement of expertise *per unit cost* for individual scaffolding is smaller than the group scaffolding. These results indicate that although targeted individual scaffolding may improve the expertise gain of a set of students more than group-based scaffolding, the former is less-economical when applied in a non-adaptive manner.

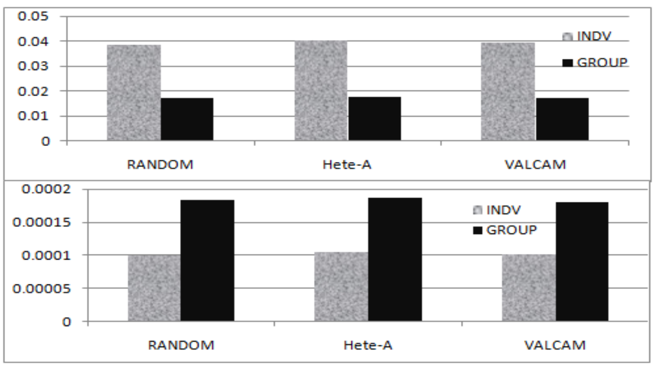

Fig. 9. (a) (top) Average expertise gain for individual and group scaffolding (b) (bottom) Average expertise gain per unit cost for individual and group scaffolding.

E. Validity and Correlation with CSCL Results

Here we validate the design of SimCoL by running several basic simulations to see whether the results verify previously published collaborative learning patterns.

Variance in Learning Rate. As reported in [5], high-ability students have higher learning rates than low-ability students because they are able to grasp, process, and internalize information received during the collaboration process. When we compared the learning rates of high/low ability learners in SimCoL, we found that the high-ability students learn at a faster rate than the low-ability students (**0.1 vs. 0.3**).

Convergence of Learning Rates. CSCL researchers [18] described the collaborative knowledge building as a cyclic process that converges to a final value. Researchers [5],[11] also described that the rate at which the students gain expertise is faster in the beginning and then slows down over time. In the beginning, the members of the groups have expertise values that are different from one another. As they collaborate with each other, the expertise values are updated and the differences among their expertise values are reduced. The total expertise gain curve shown in Fig. 10 has two properties: (1) the total expertise gain of the students converges to a final value and (2) the rate of change of the curves is higher in the beginning and slows down at the end. Furthermore, the same convergence pattern is observed when the simulation run is repeated with Hete-A group formation method. So, the knowledge gain of the students in SimCoL follows patterns described by other CSCL researchers [28-29].

Correlation with Observed CSCL Results. Here we try to compare our simulation results with CSCL results published in [22]. For this comparison, we have first mimicked a simulated environment as the CSCL classroom [22] by setting the parameters of SimCoL equal to the parameters of the CSCL classroom [22], i.e., we set: (1) the mean expertise of the students in SimCoL as 0.7, (2) number of tasks as 5 for each collaborative learning session, (3) number of students as 11, and (4) mean social relationship of the students as 0.9. Then similar to the CSCL classroom, we have simulated 4 collaborative sessions in SimCoL. Then we have calculated the correlation between the actual CSCL results and simulated results in SimCoL. Table VIII shows that for both expertise gain and social relationship change, the correlation was significant and high. However, as we have discussed in Section V-B, student attributes like motivation *may* also impact the student expertise gain which we have not collected data upon. So, this correlation can be made stronger with the consideration of those factors which is in our future plan (Section VI).

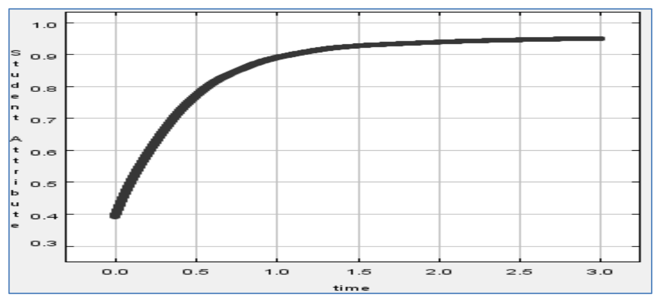

Fig. 10. Total expertise gain of students collaborating in groups formed by random group formation method.

TABLE VIII
CORRELATION BETWEEN SIMCOL AND OBSERVED CSCL[22] RESULTS

Attribute	Correlation
Student Expertise	**0.83**
Student Social Relationship	**0.97**

VI. RELATED WORKS

Table IX compares the SimCoL with other relevant simulation tools according to the considered factors in the environment. As we have described in Section II (with published learning theories), the agent-based support and group formation algorithms are two key issues uniquely discussed in SimCoL which improves the current state of the art of similar simulation tools.

TABLE IX
COMPARISON OF SIMCOL WITH OTHER EDUCATIONAL SIMULATION TOOLS

Considered Attribute/Aspect in Simulation	SimCoL	Sklar and Davies[4]	Spoelstra and Sklar[5]
Learning due to Collaborative Interactions	✓	---	✓
Student Knowledge	✓	✓	✓
Student Ability	✓	✓	✓
Student Motivation	✓	✓	✓

Student Emotion	✓	---	✓
Student Social Relationship	✓	---	---
Group Composition	✓	---	✓
Group Formation Algorithm	✓	---	---
Agent-Based Scaffolding	✓	---	---

✓ ***Considered --- Not Considered***

VII. Conclusions

The evolving domain of learning theories and CSCL systems [23] implies that a simulation environment could provide a low-cost tool to the researchers and teachers to better understand the impact of instructional approaches. In this paper, we have proposed SimCoL, an agent-based tool for simulating the collaborative learning in a CSCL system. We have described the design and implementation of the SimCoL environment and its agents using observations reported by the researchers working in the individual, peer-based and collaborative learning domains. The overall simulation results of the SimCoL environment is consistent with previously reported collaborative learning patterns. Furthermore, our results hint that the SimCoL environment allow the researchers to gain better insights into the impact of: (1) *individual* student attributes, (2) various agent-based and non-agent based group formation algorithms, (3) different types of scaffolding processes on the collaborative learning outcome of students, and (4) CSCL and collaborative learning on real classrooms in particular, and any human-computer environments where online collaborative activities take place among users with diverse behaviors.

Our future work involves gaining insights into a CSCL classroom by using a *what-for* simulation scenario. For example, for a given CSCL result, we would first match the observed outcome of SimCoL with the CSCL results by tweaking the input parameters and those *required changes* in the parameters would allow us to gain valuable insights into the environment dynamics (e.g., which of the student attributes was the dominant factor in determining the CSCL outcome) of that CSCL setting. However, for a realistic what-for simulation scenario, we need actual *measured* data regarding *all* input parameters of SimCoL. We are now working on a large-scale CSCL experiment that would involve collecting data for all critical student attributes (e.g., motivation) which will enable us to conduct the what-for simulation experiments. In future, we also plan to (1) further investigate how the time, cost, and effort invested by the students towards their groups impact the motivation, social relationship, and expertise gain by comparing the CSCL experiment data with SimCoL's simulation results and (2) we plan to compare the students' performances by comparing the RAND indexes of the simulated and real-world CSCL data.

References

[1] D. Brooks and D. Shell, "Working memory, motivation, and teacher-initiated learning," J. Sci. Edu. & Tech., vol. 15, 2006, pp. 17-30.

[2] K. Issrof and T.D. Soldato, 'Incorporating Motivation into Cooperative Learning," Proc. European Conf. AI in Education, 1996, pp. 284-290.

[3] N. Khandaker, L-K., Soh, and H. Jiang, "Student Learning and Team Formation in a Structured CSCL Environment," Proc. ICCE, 2006, pp. 185-192.

[4] E. Sklar and M. Davies, "Multiagent simulation of learning environments," Proc. AAMAS, 2005, pp. 953-959.

[5] M. Spoelstra and E. Sklar, "Using simulation to model and understand group learning," Int. Trans. Systems Science & Apps., vol. 4, 2008, pp. 76-88.

[6] DF. Shell and D.W. Brooks, "The Unified Learning Model: Implications for Learning and Teaching," 2007; http://dwb4.unl.edu/index.html.

[7] H. Ellis, et al. "Emotion, motivation, and text comprehension: the detection of contradictions in passages," J. Exp. Pysch.: General, vol. 126, 1997, pp. 131-146.

[8] A. Inaba, et al., "How Can We Form Effective Collaborative Learning Groups?," Proc. ITS, 2000, pp. 282-291.

[9] L. Vygotsky, "Mind in Society," 1978, Cambridge, MA: Harvard University Press.

[10] P. Dillenbourg. What do You Mean by Collaborative Learning?, 1999, Oxford: Elsevier.

[11] G. Stahl, "Building Collaborative Knowing: Elements of a Social Theory of CSCL, What We Know about CSCL and Implementing It in Higher Education," 2004, Norwell, MA.: Jossey-Bass Publishers.

[12] K. Kreijns, A. Paul, and W. Jochems, "Identifying the Pitfalls for Social Interaction in Computer-Supported Collaborative Learning Environments: a Review of the Research," Computers in Human Behavior, vol. 19, 2003, pp. 335-353.

[13] J. Zumbach, J. Schönemann, and P. Reimann, "Analyzing and supporting collaboration in cooperative computer-mediated communication, 2005 Conf. Computer Support for Collaborative Learning: Learning 2005: the next 10 years!,", 2005, pp. 758-767.

[14] K. Issroff and A. Jones, "Learning Technologies: Affective and Social Issues in Computer-Supported Collaborative Learning, Computers & Education," vol. 44, 2005, pp. 395-408.

[15] E. Vass, "Friendship and Collaborative Creative Writing in the Primary Classroom, J. Computer Assisted Learning," vol. 18, 2002, pp. 102-110.

[16] T. Clear and D. Kassabova, "Motivational patterns in virtual team collaboration," 7th Australasian Conf. Computing Education, 2005, pp. 51-58.

[17] J. Bruner, J. "Vygotsky: A historical and conceptual perspective, Culture, communication, and cognition: Vygotskian perspectives," 1985, Cambridge, England: Cambridge University Press, 21-34.

[18] C. Cazden, "Peekaboo as an instructional model: Discourse development at home and at school, Stanford Papers and Reports in Child Language Development," vol. 17, 1983, pp.1-19.

[19] A. Harrer, N. Malzahn, and A. Wichmann. "The remote control approach - An architecture for adaptive scripting across collaborative learning environments," J. Universal Comp. Sci., vol. 14, 2008, pp. 148-173.

[20] X. Liu, B. Doore, and L. Li, " Scaffolding Knowledge Co-Construction in Web-based Discussions through Message Labeling," Proc. Society for Inf. Tech. & Teacher Education Int. Conf., 2008, pp. 3041-3046.

[21] A. Gogoulou, A. et al. "Forming Homogeneous, Heterogeneous and Mixed Groups of Learners," Proc. ICUM, 2007, pp. 33-40.

[22] L-K. Soh, N. Khandaker, and H. Jiang, "Multiagent Coalition Formation for Computer-Supported Cooperative Learning," Proc. IAAI, 2006, pp. 1844-1851.

[23] G. Stahl, T. Anderson, and D. Suthers, "Computer-supported collaborative learning: An historical perspective," 2006; http://www.cis.drexel.edu/faculty/gerry/cscl/CSCL_ English.pdf.

[24] Repast. 2008. Recursive Porus Agent Simulation Toolkit. http://repast.sourceforge.net.

[25] J.V.D. Linden, and P. Renshaw, "Dialogic Learning," 2001, Dordrecht: Kluwer Academic Publishers.

[26] P.M. Fitts, "Perceptual-motor skill learning," New York: Academic, 2001.

[27] J.R. Anderson, "Acquisition of Cognitive Skill", Psychology Review, vol 89, 1982, pp. 369-406.

[28] W.J. Clancey, "Roles for agent assistants in field science: Understanding personal projects and collaboration",IEEE-SMC, vol. 34, no. 2, 2004, pp. 12-137.

[29] T.S. Roberts and J.M. McInnerney, "Seven problems of online group learning (and their solutions)", Educational Technology and Society, vol. 10, no. 4, 2007, pp. 257-268.