Walk before you run: using heuristic evaluation to assess a training tool prototype

Miguel J. Monasor
Lero, University of Limerick
Alarcos Research Group, University of Castilla-La Mancha
MiguelJ.Monasor@gmail.com

John Noll
Lero, The Irish Software Engineering Research Centre, University of Limerick
+353-61-202956
John.Noll@lero.ie

Aurora Vizcaíno
Alarcos Research Group, University of Castilla-La Mancha
+34 926 295300 ext. 6478
Aurora.Vizcaino@uclm.es

Mario Piattini
Alarcos Research Group, University of Castilla-La Mancha
+34 926 295300 ext. 3715
Mario.Piattini@uclm.es

Sarah Beecham
Lero, The Irish Software Engineering Research Centre, University of Limerick
+353-61-233769
Sarah.Beecham@lero.ie

ABSTRACT

Context: One of the chief challenges of Global Software Development (GSD) is for globally, culturally and linguistically diverse team members to communicate effectively. Failing to meet this challenge can lead to misunderstandings that impede project success. To prepare practitioners for remote and diverse forms of communication we have developed VENTURE: a simulation-based training platform that aims to increase cultural awareness.

Aim of study: Having developed a prototype training platform, we now assess whether VENTURE can deliver real learning and increase cultural awareness.

Method: A survey was designed based on educational theory. Using a heuristic evaluation approach, a group of potential users trialed the system and completed a proof of learning survey before and after using the platform. Survey results, and VENTURE's own automated assessment scores, were analysed and compared.

Results: Results indicate that the platform has the potential to provide useful and meaningful training for cultural awareness in GSD. A secondary result is that the assessment scheme proved useful in providing both evidence of learning and highlighting areas where VENTURE could be improved.

Conclusion: Researchers and practitioners can gain an understanding of how to evaluate, train and measure soft skills, such as communication, important to all forms of software development. The methodology applied served to adapt the environment to the needs of potential users and the actual problems of GSD.

Categories and Subject Descriptors

K.3.2 Computer and Information Science Education

General Terms

Measurement, Experimentation, Human Factors, Languages

EASE '14, May 13 - 14 2014, London, England, BC, United Kingdom

http://dx.doi.org/10.1145/2601248.2601271

Keywords

cross-cultural project; interactive learning environments; simulation; teaching/learning strategies; global software development; education

1. INTRODUCTION

During the last two decades, traditional co-located software development activities have evolved towards a geographically distributed model carried out by groups of people who do not necessarily share the same culture, skills and knowledge [1]. Global Software Development (GSD) is now commonly applied by multinational companies who aim to reduce costs, increase the speed of responding to the market, and gain access to a diverse, qualified workforce and new markets [2].

However, this transformation into a multi-site, multilingual, multicultural environment creates additional complexities beyond those already present in co-located software development [3]. As such, participants in GSD must develop new skills not necessary in co-located development [4], due to working in teams with cultural and linguistic differences.

Both universities and companies must therefore invest in training in order to effectively tackle the problems of GSD, especially those related to communication, collaboration and coordination. However, training in these soft skills is not well suited to traditional training methods (e.g. paper based, and theoretical). Ideally training should be provided in a practical context, where students are placed in realistic situations, which traditional training approaches find difficult to emulate [5].

To address this gap, we developed *VENTURE* (Virtual ENvironmenT for commUnication and collaboRativE training) [6], a training environment specifically designed to help people learn the skills needed in GSD, in a systematic and guided manner. VENTURE trains learners to recognise GSD type communication problems by means of simulation, in which learners interact with two types of Virtual Agent, both of whom play a specific role in the training process. The learner engages in a simulated instant messaging (or email) dialog with a *Virtual Colleague*, who represents a developer from a different culture from that of the learner. A *Virtual Guide* observes the dialog, correcting the learner's cultural mistakes, and providing immediate feedback. In this way, VENTURE combines theoretical learning based on the study of GSD scenarios, the use of simulations that reproduce these scenarios, and an automated

assessment of the learner's progress, to provide just-in-time training in cross-cultural interaction [7].

This paper describes a study to evaluate the effectiveness of VENTURE in providing cultural training in GSD. In this study, we applied a Heuristic Evaluation (see Section 4) to test whether this training approach increases learners' competence in communicating with people of different cultures and first languages. We also assessed whether the GSD training scenarios reflect realistic problems and situations. Potential users of VENTURE tried the training environment, then completed a survey. Heuristic Evaluation exposed gaps in our training scenarios, and helped to guide our future development efforts.

The main objective of this evaluation was to obtain feedback about the prototype chat simulator and scenario designer and assess whether the training tool has the potential to provide GSD training, and can therefore be effective in giving the student increased confidence to carry out effective communication with people of different cultures and languages. The specific sub-goals of this work are to:

- Elicit advice and feedback in order to improve VENTURE.
- Identify future training scenario designs.
- Identify improvements that it might be necessary to make in order to adapt VENTURE to such scenarios.
- Evaluate the difficulties of applying the platform in universities and companies.

Study participants reported that they found VENTURE usable and potentially effective. They viewed simulation as a good alternative to traditional training methods. They also offered suggestions for new training scenarios for both industry and academic settings. These suggestions, combined with the results from the survey and the automated assessment of the users' interactions with the Virtual Colleague, show that VENTURE has the potential to help students and practitioners to develop the skills required for effective communication in a GSD context.

The remainder of this paper is organized as follows: Section 2 explains the theoretical background of this research. Section 3 explains VENTURE. Section 4 describes the research objectives of this work and Sections 5 explains the research method followed. Section 6 summarises the results of the evaluation, which are discussed in Section 7. Section 8 explains the limitations of this study. Conclusions and future work are detailed in Section 9.

2. RELATED RESEARCH

Global software development requires teams to work together across geographic, cultural and linguistic boundaries. This diversity creates new challenges relating to team communication, collaboration, and coordination, that are not present in co-located development [1]. As a result, delays, misunderstandings, lack of trust, conflicts, and lack of team awareness, increase the cost of development [2].

The cultural barrier in GSD is often accompanied by geographical and temporal distance [8], that call for new management approaches. For example, the absence of nonverbal cues and face-to-face contact require different forms of reward structures to motivate individuals working in virtual teams [9, 10]. However, there is a significant variance in what individuals regard as a positive reward according to their culture [11]. Participants in GSD projects must therefore be aware of the cultural characteristics of others so they are not offended by certain behaviours, and are tolerant of certain attitudes.

Synchronous and asynchronous communication between co-workers in any development situation play an important part in building trust and social relationships; this is especially true in GSD [12]. It is therefore necessary to train participants to use both forms of communication effectively [13].

2.1 Cultural and linguistic differences in GSD

MacGregor et al's [14] study of GSD projects identified a set of different cultural patterns of behaviour. These patterns vary according to the cultures of the interacting participants, and can lead to conflicts, inefficiencies and misunderstandings. Consequently, these patterns have a significant influence on team performance; appropriate training may help mitigate them.

A common example of cultural conduct that can generate conflicts is the use of direct and indirect speech. People in low-context cultures tend to speak without elaboration [11], whereas explicit communication is necessary in GSD [15]. Participants from low-context cultures should therefore be trained to acquire skills and practices that enable them to be more explicit.

Cultural differences are often accompanied by differences in language skills. For example, native speakers have a natural tendency to assume facts that may result in the loss of valid information and misinterpretations that can eventually damage team relationships [16]. Practitioners must know how to communicate by following guidelines [17], such as: formulate criticism and praise carefully; avoid slang, colloquialisms, jargon, acronyms and metaphors; avoid humor and jokes that may be misinterpreted.

2.2 Teaching and training GSD

Since GSD requires frequent communication and close collaboration [11], training in these areas should promote effective teamwork in virtual teams [5]. New attitudes and competencies that must be acquired to work effectively on a GSD project [18] include the following:

- Communicating openly to generate trust [19].
- Facilitating clear communication to avoid misunderstandings.
- Specifying responsibilities and how to get things done [19].
- Detecting when attention is drifting, and bringing the members back on track [11].
- Understanding the point of view of the other participants [19].
- Negotiating effectively and avoiding conflicts [20].
- Knowledge of cultural principles, customs and language [21], [22].
- Information management, synthesis, analysis and critical reasoning skills [21].

Some universities provide training in GSD by replicating the conditions of real environments [23]. However, providing training on specific GSD skills requires reproducing accurate cultural and linguistic problems, which are difficult to systematically reproduce in educational settings [5].

2.3 Use of Simulation in Education

Simulation is often used in Software Engineering educational courses as it helps to maximize the learner's transferability of academic knowledge to real-world settings [24]. As an example, a role-based collaborative learning approach is presented by De Lucia et al. [25].

Conversational agents that interact naturally with learners have also been applied in education. For example, Veletsianos et al. [26] examines the effective deployment of conversational agents in virtual worlds from the perspective learning technologies and engineering.

Simulation and Virtual Agents seems well-suited to GSD training, since agents can play different roles and simulate people from different countries, which can help to reproduce difficult GSD situations.

3. VENTURE

VENTURE is a training platform that applies simulation, interactive drama, and Virtual Agents, to provide training focused on cultural and linguistic problems. VENTURE aims to provide appropriate training and feedback that incorporates both behavioural patterns and cultural norms. Simulations capture and reflect expert knowledge and experience, and lessons learned from the related cultural and GSD literature [6].

The architecture of VENTURE is described in [6] and depicted in Figure 1. An *e-learning platform* is the main interface for the learners, through which they have access to the learning materials.

After studying these materials, learners can participate in simulated chat and e-mail dialogs in the context of a predefined training scenario. Their performance during these simulations is automatically assessed by VENTURE, which provides access to their specific results in the *evaluation area*.

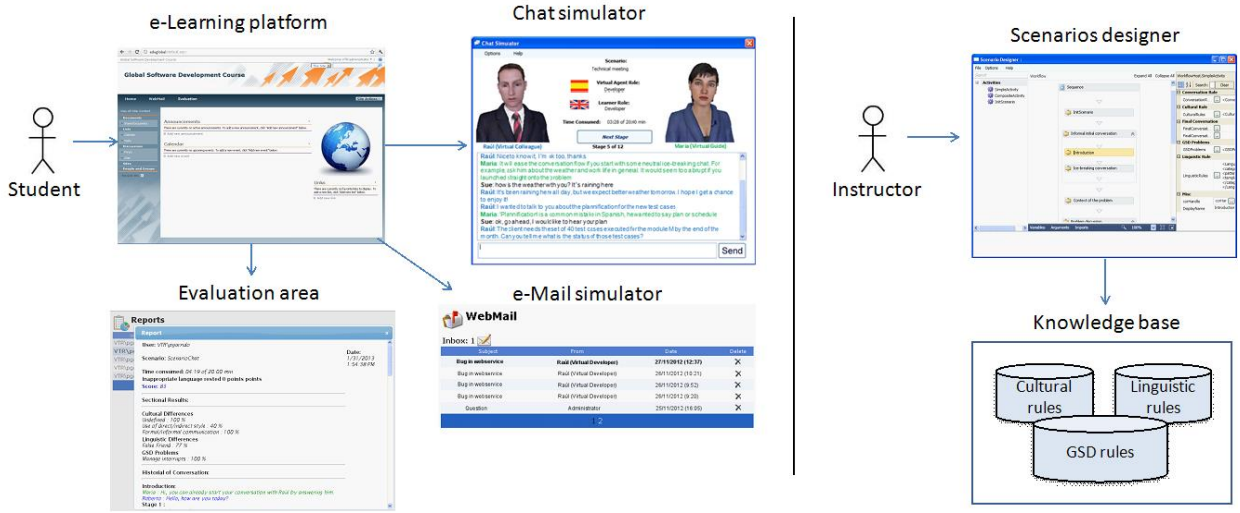

Figure 1. VENTURE's environment.

The ability for instructors to design their own scenarios, rules, chat and email simulations is provided by VENTURE's *scenarios designer*. Instructors can also access the *e-learning platform* to manage their courses and resources, assign training modules, and track the learners' progress and self-evaluation. The main components of VENTURE are described in detail in the following sub-sections. For further details go to http://global.lero.ie/venture.

3.1 Chat and e-mail simulators

Figure 2. Chat Simulator interface.

VENTURE can simulate synchronous (chat) or asynchronous (e-mail) dialogs. In the case of the synchronous interactions, the learner's task is to obtain as much information as possible within the time limit of the simulation, thus allowing additional training in time management skills. When a learner executes a simulation, VENTURE's workflow engine loads the definition of the training scenario, and presents an introduction to the scenario, after which the simulation begins. Learner's interact with Virtual Agents by means of a chatbot system [27]. The implementation of this system applies concepts of narrative drama, in the sense that interactions are made up of different phases in which specific topics are explored, and related conversational rules applied.

An example of a synchronous interaction is presented in Figure 2. In this example, an English learner (Sue) plays the role of a software developer in a multinational company.

The learner has to interact with the Virtual Colleague, who represents another developer from Spain called Raúl involved in the same project. During the simulation, VENTURE creates Raúl's answers by applying the conversational knowledge defined for each phase of the conversation. Maria, VENTURE's Virtual Guide, will advise and correct the learner in real time by applying the corresponding cultural and linguistic rules. The Virtual Agents are animated and so can also react by means of gestures.

3.2 Scenario designer

Instructors can create their own training scenarios using VENTURE's Scenario Designer, which allows the instructor to define the training scenario to include: characteristics of the

virtual agents involved, the duration of the simulation, and the title, and description of the simulation. Instructors can also add phases into the scenario. Furthermore, for each phase, specific rules can be assigned by introducing patterns that will be detected during the simulation. Instructors can also design templates that trigger the Virtual Guide to provide advice when a rule has been fired during an interaction with the learner.

To further facilitate the definition of the training scenarios, instructors can retrieve cultural and linguistic rules from a database, which contains a set of predefined rules organized by their type. Similarly, when the instructor creates a new rule, it can be added to the database and reused in the future.

3.3 Automated assessment

VENTURE automatically assesses learner's progress by considering the rules fired during the interactions. Figure 3 shows a report indicating a student's score. Students start with a perfect score of 100; if no errors are made throughout the interactions, their final score will be 100. Each rule contained in the scenario has an associated severity mark, indicating the score that will be subtracted if the user fires such a rule.

The report includes information about the date and duration of the simulation, along with the number of times that the user gets stuck and has to skip a stage. It also includes a final score and sectional results, detailing the scores obtained for each GSD skill that was addressed in the training scenario.

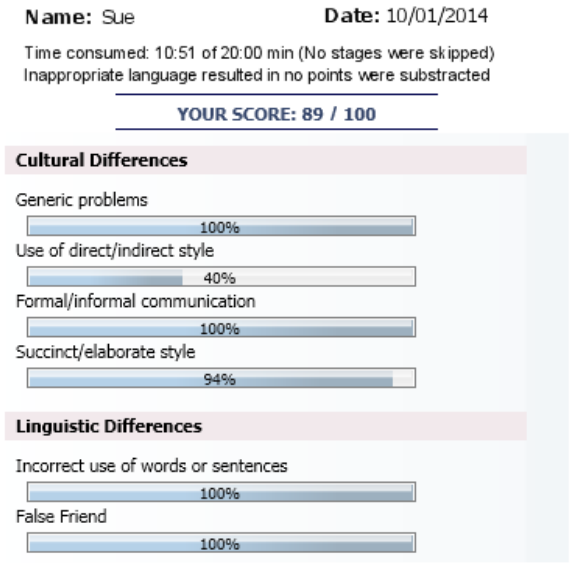

Figure 3. Automatic assessment.

4. RESEARCH METHOD

The research objective of the study described in this paper is to evaluate the effectiveness of VENTURE in providing cultural training in GSD. The need for this evaluation arises from the complexity of GSD. Two aspects of training solution need to be evaluated: firstly, we need to know whether the simulated training environment is suited to real learning, and secondly, whether the application and content, based on rules and scenarios can deliver useful learning in GSD-related skills. As a consequence, the opinion and feedback of experts and potential users is needed before releasing VENTURE into real educational settings.

The current study is part of a larger research approach depicted in Figure 4, which began with a set of systematic literature reviews to identify the issues and gaps in current approaches to cultural awareness training in GSD [2]. The next step consisted of designing VENTURE´s simulation platform and example training scenarios. A preliminary Expert Evaluation was conducted in order to get feedback on the initial architecture and proof of concept [28] (innovation and develop phase). This feedback informed the development of the Version one (phase 3 in Figure 4), the first working prototype of VENTURE that can be used by actual learners.

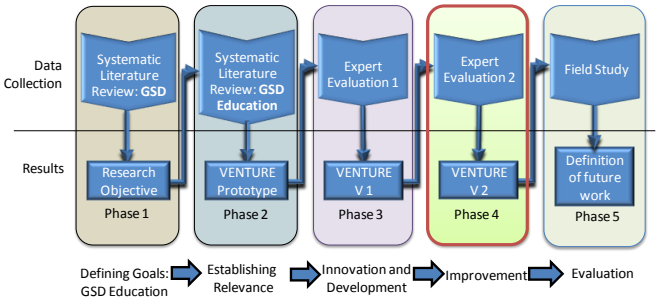

Figure 4. Research method.

Version one was used for Expert Evaluation 2 (Figure 4, phase 4), the Heuristic Evaluation described in this paper. The feedback from this evaluation, will, in turn, inform the development of VENTURE Version two by incorporating suggestions in design and content that experts recommend in this study. Version two will then be used in an extensive field study designed to assess VENTURE's effectiveness as a learning tool.

We chose Heuristic Evaluation [29] as the method for the second Expert Evaluation in Figure 4. The Heuristic Evaluation method involves a small but varied group of experts representing different relevant fields of expertise. Using Heuristic Evaluation has the following advantages [30]: the application of recognized and accepted principles; intuitiveness; usability early in the development process; effective identification of major and minor problems; rapidity; and usability.

A well designed Heuristic Evaluation has the advantage that a small number of participants can provide both a deep and broad assessment. Participants should represent experts and potential users, and come from varying disciplines [31]. Although there is no consensus as to the optimal number of experts required to evaluate a system, Hwang and Salvendy [32] indicate the ideal number of participants for this form of evaluation is between eight and twelve.

4.1 Research setting

As shown in Table 1, study participants were specifically selected from complementary disciplines: research (and teaching), and GSD practice. The sample was opportunistic in that participants were recruited through two main sources; firstly delegates were approached at the annual international conference for global software engineering (ICGSE) that comprises both practitioners and researchers, and secondly post graduates, doctorates, and professors who were researching in software engineering and based at the University of Limerick where solicited. Each participant was selected based on either their knowledge of GSD, or their knowledge of training, or their knowledge of tool building. The sample therefore was stratified and opportunistic.

Our sample of eighteen participants is slightly over the recommended number because we wanted to include three user groups.

Table 1. Roles of the participants in our Heuristic evaluation

Type of Participants	Number of participants	Percentage
Researchers	7	39%
Practitioners	3	17%
Practitioners/Researchers	8	44%
Total	18	100%

As depicted in Table 1, eight of the participants (44%) classed themselves as practitioners and researchers, as they were involved in both fields. Moreover, three participants (17%) were purely practitioners and the other seven (39%) were purely researchers. Of the researchers, two are university professors who teach subjects directly related to GSD.

Participants represented thirteen different countries: Argentina, Brazil, Finland, Germany, India, Ireland, Italy, New Zealand, Nigeria, Netherlands, Nepal, Pakistan and USA. The participants were selected by considering their availability and their experience in GSD. The average experience in GSD of the participants was 8 years.

Table 2 summarises the size of the practitioner and practitioner/researcher participant's company, the size of the IT department, and the number of countries usually involved in GSD activities. The companies represented are mostly large multi-national organisations, typical in GSD.

Table 2. Company characteristics of practitioner participants

Company size	IT dept size	Countries involved in GSD
100000	10000	5
150000	130000	10
4500	3500	9
4500	3500	6
18000	1000	20
100	95	1

The practitioner respondents included project managers, general managers, researcher (in two cases), marketing managers and quality and process managers; the headquarters of the companies were located in four countries: India, Finland, Brazil and Germany, providing a good east/west mix of experience.

4.2 Survey procedure

For this evaluation each participant used both the Chat Simulator and the Scenario Designer. Each participant spent approximately 20 minutes using VENTURE, and then 10 minutes completing a short questionnaire containing both open- and closed-ended questions (see Appendix A).

Prior to the actual trial, each participant was given a verbal explanation of the objectives of the evaluation. Then, the participant was given some background information explaining that during the simulation the participant would play the role of software developer, and that he or she would be required to interact with a virtual Spanish developer.

Participants then executed a short training scenario using the Chat Simulator on their own. During the execution of the scenario, VENTURE produced its own automated assessment of the participant's performance regarding the number of times a given rule was fired (which we later analysed to see where the learner made mistakes). During the simulation, the Virtual Colleague (simulating a native Spanish speaker) made some typical Spanish mistakes when speaking English. The Virtual Guide explained these problems to the participants and also corrected some of their incorrect interactions. For example, they were advised when they were too direct in addressing a problem.

On completion of the Chat simulation, participants were shown VENTURE's Scenario Designer. They were shown how the scenarios were defined by dragging and dropping actions, and how the rules were added. Participants examined the definition of the scenario that they had executed during the Chat simulation, including its settings and the rules that they had fired. Finally, the participants completed the questionnaire and provided their feedback on both the training and design elements of VENTURE. This included questions on user experience and future improvements that complemented the results obtained by the automated assessments that were also analysed.

5. RESULTS

This section reports on the results collected from our survey and from scores generated during the execution of the scenarios.

5.1 Quantitative Results

We analysed two sets of quantitative results; firstly those generated by VENTURE's automated assessment facility from data collected during the execution of the training scenarios, and secondly results from the closed-ended survey questions that required the participant to respond using an ordinal Likert scale.

5.1.1 Rules fired

Table 3. Results of the evaluations.

Parti-cipants	Rules fired	Responses to questions (Q7.1-Q7.8)							
		7.1	7.2	7.3	7.4	7.5	7.6	7.7	7.8
1	2	5	4	3	5	3	5	5	5
2	1	5	3	3	2	4	3	4	5
3	2	4	3	4	4	3	3	5	4
4	2	4	2	3	3	3	4	4	5
5	2	5	3	3	4	2	4	4	3
6	2	4	3	3	3	4	3	3	3
7	0	5	3	4	NA	5	4	4	3
8	2	5	4	4	3	4	3	5	4
9	1	5	5	5	4	3	5	5	5
10	0	5	3	4	4	4	4	4	4
11	0	5	4	3	3	2	5	5	4
12	0	3	4	3	3	4	4	4	4
13	3	4	3	3	5	5	4	5	4
14	1	5	3	3	4	4	4	5	5
15	1	4	4	2	3	4	4	4	2
16	0	4	4	4	3	3	3	3	4
17	4	5	4	4	4	4	5	5	4
18	3	4	4	4	4	3	5	4	3
Median	**1.5**	5	3.5	3	4	4	4	4	4

Recall the example in Figure 2 in which the Virtual Guide provides feedback after detecting that the learner has used a wrong word in a certain context of the interaction. During the execution of the scenario, most participants made some interaction mistakes as detected by VENTURE, resulting in an intervention from the Virtual Guide to provide feedback. This

feedback is triggered when a learner's mistake causes one of VENTURE's cultural interaction rules to fire.

The aggregated results of these rule firings from the eighteen trials are summarized in the "Rules fired" column in Table 3. For example, participant 1 made two mistakes, causing two cultural interaction rules to fire.

The results in Table 3 show that only four of eighteen participants managed to conduct the interaction without making an error. This suggests that even an expert would be challenged by the scenario's level of difficulty. Most of the participants (twelve) made at least one mistake, causing at least one cultural rule to fire (as depicted in Table 3). No linguistic rules were fired owing to the participants' high level of English.

5.1.2 Evaluation questionnaire

Table 3 also shows the participants' responses to the closed-ended questions of the evaluation questionnaire, which they scored from 1 (strongly disagree) to 5 (strongly agree). The Likert scale questions (which are a part of the larger questionnaire detailed in Appendix A) were designed to assess participants' impressions of both VENTURE's Chat Simulator and Scenario Designer, and are listed here:

Q7.1: The Chat Simulator is easy to use
Q7.2: The Chat Simulator can be used to train people to recognise linguistic differences in GSD
Q7.3: The Chat Simulator can be used to train people to recognise cultural differences in GSD
Q7.4: The Chat Simulator is effective in correcting learner's mistakes (e.g. use of ambiguous language or inappropriate behaviour)
Q7.5: The information in the final report provides learners with useful guidance – i.e. they can, given this information, reflect on how they can improve their linguistic and cultural communication.
Q7.6: The Designer can highlight typical cultural and linguistic mistakes in GSD
Q7.7: I enjoyed using the Chat Simulator
Q7.8: Using a Chat Simulator is a good way to train individuals (as a concept)

Median values of responses to each question shown at the bottom of Table 3 indicate that the participants' assessment of VENTURE range from 3 (neutral) and to 5 (strongly agree); only a few (5) individual responses were below 3. This means that the overall perception of VENTURE is positive.

The participants gave VENTURE high marks for ease of use (Q7.1), and most of them enjoyed the experience (Q7.7). Participants were also favorable of the feedback provided by VENTURE's Virtual Guide (Q7.4) and assessment report (Q7.5). Further, they considered the scenario designer a potentially useful tool for designing cultural training scenarios (Q7.6). With few exceptions, participants felt that VENTURE is an effective approach to cultural training (Q7.8).

Participants were not quite as confident of VENTURE's overall potential for effective cultural training: the median response to Q7.2 regarding the platform's potential for helping learners to recognize linguistic differences was slightly above neutral (3.5), and the response to Q7.3 regarding VENTURE's potential for helping learners to recognize cultural differences was neutral (3).

Questions 7.2, 7.3 and 7.4, 7.5 respectively may appear similar, but serve a different purpose. Questions 7.2 and 7.3 were designed to elicit responses on the perceived usefulness of the platform to generate valuable learning outcomes. Questions 7.4 and 7.5 on the other hand, are aimed at discovering the effectiveness of the training scenarios (the stories used to teach new concepts). Responses to the second group of answers were more positive than the answers to the first group. This is not surprising since although participants were able to enhance their knowledge by using VENTURE, the participants observed that the robustness should be improved before application could be used in educational settings. The answers to the survey's open-ended questions, which are discussed in the next section, give some insight into why participants responded this way.

In summary, the results of the quantitative results show that VENTURE is able to provide feedback that allowed participants to correct their interaction based on GSD rules. Moreover the experience of the participants was enjoyable.

5.2 Qualitative results

The remaining qualitative questions (1-6) and answers that were gathered through the questionnaire in Appendix A are summarized in this section.

Q1. Do you think this training method can be effective in the accurate training of specific cultural and linguistic differences in GSD?

Most of the ICGSE respondents who fired cultural rules answered positively to this question. For example, one participant wrote, "It seems to have the potential to be flexible enough to cover many aspects". Another remarked that the Chat Simulator is interesting as complementary material although it could not be the core of the training method. One respondent who did not fire any cultural rules stated: "It is a good start but it would need more rules and enhancements", although this respondent understood that the scenario is very limited in time, and a real scenario would have more semantic depth and deal with realistic problems.

The opinions of one of the researchers was quite similar: "I like the way Maria (Virtual Guide) corrects what is going wrong… when I was too direct she pulled me up straight away. She was also quick to point out Raul's (Virtual Colleague) incorrect use of words".

Q2. Do you think it would be feasible to train students/members of your university/company by applying this environment?

Most of the respondents agreed VENTURE would be a feasible approach to train learners or members of their university or company. In the words of one of participant: "I could see it being used in classes by learners and it could help to train them in cross cultural inequity".

One ICGSE respondent thought the simulator would be more suited to company training than university education. The main reason for this is that he found it specific to the training of certain problems that can appear in the specific settings of each company. Another practitioner thought that it would be more useful if this training could be conducted with real learners in real time. However one of the aims of this type of training tool is that it tries to anticipate and prepare practitioners before any real interaction.

Q3. What problems did you experience while using the Chat Simulator?

An analysis of responses to survey question 3 indicates that most participants did not have any problems using the system during the simulation. One participant admitted that sometimes he did not know how to continue, but there were no great problems as he had the possibility to skip the phase rather than stall the training.

Two participants attempted to challenge the tool and found some inconsistencies when they tried to be rude by using offensive words that were not taken into consideration in the scenario definition. Sometimes they received good advice from the Virtual Colleague, such as "Avoid using expletives. This type of language can be offensive". Other times, VENTURE failed to recognize some offensive words and therefore did not give feedback.

One of them also tried to be very direct in communication, but during a phase of the conversation in which the scenario did not expect such a direct approach, because it was out of context. One of the users also expected to receive answers to questions that he posed which were not considered in the scenario definition.

Another participant pointed out that the special words or expressions used to say the same thing can vary from one person to another and it is difficult to handle all of them.

Q4. Can you suggest any point for improvement or new features that you would like to be implemented in the simulator?

Two respondents who had interacted in a manner that was not taken into account in the design of the scenario, proposed automating the recognition of synonyms by using a thesaurus, and the recognition of language abbreviations was also suggested. For example, two different participants used the abbreviation "ws" to refer to "web service", however, the scenario definition did not expect this abbreviation.

A respondent suggested that when a rule is fired, more context information could be shown. For example, "when the Virtual Guide says that a particular Friday is a holiday in Spain", she could give a list of public holidays in Spain that could let me learn something about Spain".

One of the respondents proposed allowing the learner's mistakes to be run through again at the end of the simulation, and replaying them as a means to reinforce learning; he felt this would be especially applicable to university courses. This participant also suggested a global report that groups the common pitfalls committed by all the learners in the class this would be useful for the instructor in order to explain the most common problems that the learners in that class have.

One respondent suggested providing more background information in the summary report that VENTURE generates at the end of a session, in order to reinforce learning, for example by explaining typical characteristics of the culture of the Virtual Colleague or giving more information about why certain rules were fired.

Q5. Upon considering the training scenario in which you have participated, which other training scenarios do you think it would be interesting to design for the training of specific GSD problems or skills?

The participants suggested the following kinds of training scenarios in answer to Question 5 of our survey:

- The coordination of meetings to show differences in time perception.
- Setting deadlines to train differences in perception of pressure or hierarchy.
- Reaching the day of a deadline.
- Starting a project (introducing people).
- Dealing with a crisis. Dealing with serious problems in a project and conflict resolution. E.g. how to say that the due date will not be met.
- Discussions between people with different competence levels. E.g. experts and new comers.
- Notifying a failed acceptance test.
- Querying an implemented feature that does not map onto a requirement as expected.
- Obtaining information about specific technologies from experts.

Q6. Can you suggest any point for improvement or new features that you would like to be implemented in the Designer?

One of the respondents who challenged the Chat Simulator by using inappropriate language, suggested including a protocol in the Designer that could be used to detect inappropriate interactions during the simulations. The possibility of providing multiple paths in the timeline of the simulation was also suggested. Students could therefore choose the course of the interaction by means of their answers, thus making the conversation more dynamic, from the point of view of one of the respondents. Other minor ideas, such as increasing the number of avatars in order to cover a wider range of cultures in the simulations, were also suggested.

6. DISCUSSION

We set out to assess the effectiveness of VENTURE to provide cultural training in GSD. In this section we discuss how our empirical evaluation helps in this assessment.

For ease of analysis, we divide the discussion into identifying strengths and weaknesses of VENTURE. Particularly, we examine technical aspects (such as usability, development and architecture), and content aspects (such as the underlying depth of the training deployed, accuracy of cultural rules and usefulness of the scenarios).

We complete this section with a summary of improvement ideas, and educational outcomes.

6.1 Strengths of VENTURE

For the most part participants managed to navigate through the training scenario without intervention from the VENTURE researchers. Also in each case, VENTURE generated an assessment that reflected the scores accurately of each participant. This indicated that the usability and technical application is at least at a standard suitable for a prototype, and did not detract from the focus of the evaluation, which was to test whether the tool can be used to impart learning in culture and related GSD interactions.

The researcher participants liked the idea that VENTURE can provide independent and customized training focused on specific problems in GSD. In their words: "When a developer confronts a

global project for the first time, he can suffer from stress and fear of failure … being able to practice beforehand and learning how to interact can reduce this problem during the initial stages of the project".

In terms of learner experience, both the practitioners and researchers felt that the use of the Chat Simulator is similar to any other chat application. Therefore they found the environment familiar and easy to use. The use of Virtual Agents is seen as a good option to simulate GSD settings and provide learners with feedback. One practitioner indicated that when interacting with the Virtual Colleague, learners are not going to react in the exactly same way that they do with real people, but agreed that Virtual Agents are perfectly valid for teaching purposes.

The complete learning environment was valued as it minimizes the instructor's workload as VENTURE reduces time required to organize courses, involve experts, assess learners, compared with traditional methods. The flexibility of the training to be tailored to a specific culture's needs was seen as one of the tool's strong points. Participants felt it would be feasible to apply the tool in educational environments as well as industry training. In support of this idea, they suggested a broad set of training scenarios.

6.2 Weaknesses of VENTURE

Since the participants were aware they were evaluating a prototype, they found very little to criticize in terms of the usability or technical application. The weak points reported were mainly concerned with VENTURE's content. To enable training in a wide variety of skills and to tailor that training to the individual learner's needs would require extending the current number of training scenarios. Also, it would require more rules. Another problem was that the time required to create new training scenarios should be as short as possible, which would require a larger set of rules in the databases.

Finally, some participants also pointed out that the special words or expressions (used in scenarios to say the same thing) can vary from one person to another, as well as from one culture to another. They recognized that it would be difficult to handle all of these since they are context dependent (i.e. a thesaurus or library of synonyms would not be sufficient to capture the semantics). However, extending this part of the knowledge base should be considered to improve the robustness of VENTURE.

6.3 Ideas for improvement

After using the environment, participants suggested the following points, which will form part of the improvement effort:

Technical improvements

- Improve the usability of the Scenario Designer by automating some of the tasks related to the management of the rule database.
- Implement the automatic recognition of synonyms and abbreviations during the interactions.
- Support for the automatic detection of inappropriate or offensive interactions.

Content improvements

- Improvement of the final report in order to provide more detailed information about the specific problems that the learner had during the simulation.
- Implement the database of cultural and linguistic problems database, in addition to the e-mail simulator.
- Create a set of realistic training scenarios by considering realistic GSD problems.

Include a library of synonyms based on similar expressions as well as special words that are context dependent.

6.4 Educational outcomes

The analysis of the information gathered in the evaluation process resulted in new ideas from a training and learning perspective. One of the advantages of a simulated environment is that can include just-in-time features [7]. VENTURE takes advantage of this by automatically generating a final report immediately after the execution of a training scenario. The report shows where the learner had problems in the interaction (for example was too direct in the conversation). This report can be extended to include additional material. For example, videos and advice relating to the skills that the learners need to improve can be provided based on their performance.

VENTURE provides independent assessment; in this way the learner does not need to expose their weaknesses publicly. Assessment is an important success factor [33] and can motivate the learner to improve. VENTURE aims to provide accurate assessment and to this end, future work includes defining an accurate procedure for the automatic evaluations and learners' self-assessments. The objective is not only to evaluate the learners' performance but also to determine to what extend the learners improved their GSD skills.

The research literature is rich in generalisations, and the GSD research falls into this category. This is of course useful when trying to understand general trends. However for training purposes, and designing scenarios, we need highly context specific examples of problems and solutions. Our related work is therefore to derive patterns (reusable, implementable solutions) from empirical research and practitioner stories and experience reports. Therefore, with the objective of facilitating the generation of a rich set of training scenarios, we have developed a community website (http://global.lero.ie/community). Researchers and practitioners with GSD experience can now collaborate by contributing GSD patterns and scenarios. In this way realistic GSD scenarios will be gathered in a way that can be used to design new training scenarios that are incorporated into the VENTURE training platform.

7. LIMITATIONS

With regard to construct validity, the two evaluations conducted in this research were focused solely on the Chat Simulator and the Scenario Designer. Future evaluations are planned to test asynchronous interactions.

The training scenario used in the evaluation did not consider specific problems of the learner's culture, since it was oriented towards participants at an international conference in which multiple cultures were involved. Another limitation of this training scenario is related to the time constraints, as we needed to evaluate the tool without disturbing the participants too much.

With regards to internal validity, since the participants handed their responses directly to the authors of the study, we are aware there may be some bias in how participants answered the questions. However, we also have the automated assessment, generated by VENTURE that does not require any intervention from the authors or the learners.

8. CONCLUSIONS AND FUTURE WORK

In this paper we have presented a new form of cultural training tool. VENTURE applies simulation to GSD training in order to

provide a means to systematically reproduce GSD scenarios which are difficult to emulate in traditional training methods. We have applied the idea of interactive dialogues using embodied agents with which learners can discover the behavior and characteristics of people from other cultures. Being able to interact effectively with people from different cultures is particularly important in GSD [34].

In this paper, we presented an overview of VENTURE and its objectives which is to provide training in culture, especially relating to GSD settings. We detail an evaluation we conducted with 18 GSD experts coming from both industry and academia. In our evaluation we focused on whether VENTURE has the potential to improve practitioners' interactions with colleagues from different cultures. Our evaluation was based on a survey that included qualitative and quantitative questions. We also looked at the automatic reports that were generated after each participant interacted with the platform. In this way we gathered rich data that can inform us of the platform's strengths and weaknesses, and how well it meets its objectives.

Results of the evaluation indicate that VENTURE is easy and fun to use, and has the potential to be an effective tool for training in effective communication in a multi-cultural setting. However, the participants responses and suggestions make it clear that, to achieve this potential, VENTURE must be enhanced with a richer set of training scenarios, and deeper knowledge of linguistic and cultural

Heuristic Evaluation proved to be an excellent approach for evaluating the platform. The diversity of the population participating in the evaluation, comprising researchers, practitioners, and teachers, provided diverse opinions focused on different aspects of the VENTURE approach, such as the user interface, its usability, effectiveness for training in cultural and linguistic problems in GSD, and its applicability in university classes and companies. The analysis of the results suggests that the main objectives of the tool may be fulfilled as new scenarios are added and existing scenarios are elaborated.

As such, we advocate conducting Heuristic Evaluations of new approaches involving software tool development early in development lifecycle (as soon as a working prototype is available). Heuristic Evaluations do not require large numbers of participants providing you ensure that amongst them you include your core users, and related experts. In this way major gaps can be identified that need to be addressed before a full blown, resource hungry field evaluation.

9. ACKNOWLEDGMENTS

This work was supported, in part, by Science Foundation Ireland grant 10/CE/I1855 to Lero - the Irish Software Engineering Research Centre (www.lero.ie). It has also been funded by the GEODAS-BC project (Ministerio de Economía y Competitividad and Fondo Europeo de Desarrollo Regional FEDER, TIN2012-37493-C03-01). It is also supported by GLOBALIA (PEII11-0291-5274), Consejería de Educación y Ciencia, Junta de Comunidades de Castilla-La Mancha.

10. REFERENCES

[1] Šmite, D., C. Wohlin, T. Gorschek, and R. Feldt, *Empirical evidence in global software engineering: a systematic review.* Empirical Software Engineering, 2010. **15**(1): p. 91-118.

[2] Monasor, M.J., M. Piattini, and A. Vizcaíno, *Challenges and Improvements in Distributed Software Development: A Systematic Review.* Advances in Software Engineering, 2009: p. 1-16.

[3] Herbsleb, J.D., A. Mockus, T.A. Finholt, and R.E. Grinter, *An empirical study of global software development: distance and speed*, in *Proceedings of the 23rd International Conference on Software Engineering*. 2001, IEEE Computer Society: Toronto, Ontario, Canada. p. 81-90.

[4] Strang, K.D., *Collaborative synergy and leadership in e-business*, in *Handbook of Research on Electronic Collaboration and Organizational Synergy*, J.S. L.Wilson, Editor. 2008, IGI Global. p. 409-434.

[5] Monasor, M.J., A. Vizcaíno, M. Piattini, and I. Caballero, *Preparing students and engineers for Global Software Development: A Systematic Review*, in *International Conference on Global Software Development (ICGSE)*. 2010, IEEE CS: Princeton, NJ, USA. p. 177-186.

[6] Monasor, M.J., A. Vizcaíno, and M. Piattini, *Cultural and linguistic problems in GSD: a simulator to train engineers in these issues.* Journal of Software Maintenance and Evolution: Research and Practice (Special Issue on Global Software Engineering), 2011. **24**(6): p. 707-717.

[7] Jones, M.J., *Just-in-time training.* Advances in Developing Human Resources, 2001. **3**(4): p. 480-487.

[8] Casey, V. and I. Richardson. *Project Management within Virtual Software Teams*. in *International Conference on Global Software Engineering (ICGSE'06)* 2006. Florianópolis, Brazil.

[9] Jay F. Nunamaker, J., B.A. Reinig, and R.O. Briggs, *Principles for effective virtual teamwork.* Commun. ACM, 2009. **52**(4): p. 113-117.

[10] Beecham, S., *Motivating Software Engineers working in Virtual Teams across the Globe*, in *Software Project Management in a Changing World*, G. Ruhe and C. Wohlin, Editors. to appear 2014, Springer.

[11] Niederman, F. and F.B. Tan, *Managing global IT teams: considering cultural dynamics.* Commun. ACM, 2011. **54**(4): p. 24-27.

[12] Bly, S.A., S.R. Harrison, and S. Irwin, *Media spaces: bringing people together in a video, audio, and computing environment.* Commun. ACM, 1993. **36**(1): p. 28-46.

[13] Dittrich, Y. and R. Giuffrida, *Exploring the Role of Instant Messaging in a Global Software Development Project*, in *International Conference on Global Software Engineering (ICGSE)*. 2011: Helsinki, Finland. p. 103-112.

[14] MacGregor, E., Y. Hsieh, and P. Kruchten, *Cultural patterns in software process mishaps: incidents in global projects*, in *Workshop on Human and social factors of software engineering*. 2005, ACM: St. Louis, Missouri. p. 1-5.

[15] Kazanskaya, S. *Human Behaviors in distributed development environments: Service provider's experience (Arcadia)*. in *IEEE 6th International Conference on Global Software Engineering (ICGSE'11)*. 2011. Helsinki, Finland: IEEE

[16] Parvathanathan, K., A. Chakrabarti, P.P. Patil, S. Sen, N. Sharma, and Y. Johng, *Global Development and Delivery in*

Practice: Experiences of the IBM Rational India Lab. 2007: IBM Press.

[17] Shachaf, P., *Cultural diversity and information and communication technology impacts on global virtual teams: An exploratory study.* Inf. Manage., 2008. **45**(2): p. 131-142.

[18] Honig, W.L. and T. Prasad, *A classroom outsourcing experience for software engineering learning*, in *Proceedings of the 12th annual SIGCSE conference on Innovation and technology in computer science education*. 2007, ACM: Dundee, Scotland. p. 181-185.

[19] Kraut, R.E. and L.A. Streeter, *Coordination in software development.* Commun. ACM, 1995. **38**(3): p. 69-81.

[20] Damian, D., A. Hadwin, and B. Al-Ani, *Instructional design and assessment strategies for teaching global software development: a framework*, in *Proceedings of the 28th international conference on Software engineering*. 2006, ACM: Shanghai, China. p. 685-690.

[21] Richardson, I., S. Moore, D. Paulish, V. Casey, and D. Zage, *Globalizing Software Development in the Local Classroom*, in *Proceedings of the 20th Conference on Software Engineering Education & Training*. 2007, IEEE Computer Society. p. 64-71.

[22] Sampaio, J. and A. Moniz *Assessing Human and Technological Dimensions in Virtual Team's Operational Competences*. 2007.

[23] Deitersy, C., et al., *GloSE-Lab: Teaching Global Software Engineering*, in *International Conference on Global Software Engineering (ICGSE)*. 2011: Helsinki, Findland. p. 156-160.

[24] Deshpande, A.A. and S.H. Huang, *Simulation games in engineering education: A state-of-the-art review.* Computer Applications in Engineering Education. **19**(3): p. 399-410.

[25] Jin, S.A.A., *Leveraging avatars in 3D virtual environments (Second Life) for interactive learning: the moderating role of the behavioral activation system vs. behavioral inhibition system and the mediating role of enjoyment.* Interactive Learning Environments, 2011. **19**(5): p. 467-486.

[26] Veletsianos, G., R. Heller, S. Overmyer, and M. Procter, *Conversational agents in virtual worlds: Bridging disciplines.* British Journal of Educational Technology, 2009. **41**(1): p. 123-140.

[27] Wallace, R.S., *The Anatomy of A.L.I.C.E*, in *Parsing the Turing Test*, S. Netherlands, Editor. 2008. p. 181-210.

[28] Monasor, M.J., A. Vizcaíno, M. Piattini, J. Noll, and S. Beecham, *Simulating Global Software Development processes for use in Education: A Feasibility Study*, in *20th European Conference, EuroSPI*, F. McCaffery, R.V. O´Connor, and R. Messnarz, Editors. 2013, Springer: Dundalk, Ireland. p. 36-47.

[29] Beecham, S., N. Carroll, and J. Noll, *A Decision Support System for Global Team Management: Expert Evaluation*, in *REMIDI - International Workshop on Tool Support Development and Management in Distributed Software Projects co-located with 7th IEEE International Conference on Global Software Engineering (ICGSE)*. 2012, IEEE Computer Society: Porto Alegre, Brazil. p. 12 - 17.

[30] Holzinger, A., *Usability engineering methods for software developers.* Commun. ACM, 2005. **48**(1): p. 71-74.

[31] Nielsen, J., *Enhancing the explanatory power of usability heuristics*, in *Proceedings of the SIGCHI conference on Human factors in computing systems: celebrating interdependence*. 1994: Boston, MA, USA. p. 152-158.

[32] Hwang, W. and G. Salvendy, *Number of people required for usability evaluation: the 10±2 rule.* Commun. ACM. **53**(5): p. 130-133.

[33] Budimac, Z., Z. Putnik, M. Ivanović, K. Bothe, and K. Schuetzler, *On the assessment and self-assessment in a students teamwork based course on software engineering.* Computer Applications in Engineering Education, 2009. **19**(1): p. 1-9.

[34] Min, Q., Z. Liu, and S. Ji. *Communication Effectiveness in Global Virtual Teams: A Case Study of Software Outsourcing Industry in China*. *43rd Hawaii International Conference on System Sciences*. 2010. IEEE CS.

APPENDIX A: Questionnaire used in the evaluation

Q1. Do you think this training method can be effective in the accurate training of specific cultural and linguistic differences in GSD?

Q2. Do you think it would be feasible to train students/members of your university/company by applying this environment?

Q3. What problems did you experience while using the Chat Simulator?

Q4. Can you suggest any point for improvement or new features that you would like to be implemented in the simulator?

Q5. Upon considering the training scenario in which you have participated, which other training scenarios do you think it would be interesting to design for the training of specific GSD problems or skills?

Q6. Can you suggest any point for improvement or new features that you would like to be implemented in the Designer?

Q7. Please, indicate your opinion as regards the following points: 1 (strongly disagree), 5 (strongly agree):

- Q7.1: The Chat Simulator is easy to use
- Q7.2: The Chat Simulator can be used to train people to recognise linguistic differences in GSD
- Q7.3: The Chat Simulator can be used to train people to recognise cultural differences in GSD
- Q7.4: The Chat Simulator is effective in correcting learner's mistakes (e.g. use of ambiguous language or inappropriate behaviour)
- Q7.5: The information in the final report provides learners with useful guidance – i.e. they can, given this information, reflect on how they can improve their linguistic and cultural communication.
- Q7.6: The Designer can highlight typical cultural and linguistic mistakes in GSD
- Q7.7: I enjoyed using the Chat Simulator
- Q7.8: Using a Chat Simulator is a good way to train individuals (as a concept

Q8. Nationality, Q9. Experience, Q10. Position

For practitioners only

Q11. How many years have you worked in the Software Industry?

Q15.What is your current role?, Q13. Size of company?, Q14. Size of IT dept?, Q15. No. of countries involved in GSD