Evaluation of a simulation-based platform for training Global Software Development

Miguel J. Monasor
Lero, The Irish Software Engineering Research Centre, University of Limerick
MiguelJ.Monasor@gmail.com

Aurora Vizcaíno
Alarcos Research Group, University of Castilla-La Mancha
Aurora.Vizcaino@uclm.es

Mario Piattini
Alarcos Research Group, University of Castilla-La Mancha
Mario.Piattini@uclm.es

John Noll
Lero, The Irish Software Engineering Research Centre, University of Limerick
John.Noll@lero.ie

Sarah Beecham
Lero, The Irish Software Engineering Research Centre, University of Limerick
Sarah.Beecham@lero.ie

***Abstract*— VENTURE is a simulation-based training platform that, by using Virtual Agents, supports students to overcome some communication problems that occur in Global Software Development (GSD). This paper reports on an evaluation based on a prototype version. The objective is to determine VENTURE´s applicability in educational settings to train participants to develop the skills needed in GSD.**

Eighteen experts including researchers and practitioners used the tool and answered a questionnaire. They were generally positive suggesting that the approach could help to train practitioners in industrial settings. Their feedback helped to strengthen some aspects of the environment to make it more GSD focused and more suitable for real training environments.

The general outcome is that VENTURE has the potential to provide useful and meaningful scenarios in which to train students and practitioners in GSD. One of the main challenges is to gather a broad knowledge base for providing a complete GSD training.

Keywords- global software development; distributed software development; education, simulation, learning

I. INTRODUCTION

During the last two decades, traditional co-located software development activities have evolved towards a geographically distributed model [1] carried out by people with different cultures, skills and knowledge. Global Software Development (GSD) [2] is now commonly applied by multinational companies who need their personnel to be qualified for the new challenges that appear in this setting.

However, traditional GSD training approaches have difficulties in reproducing realistic settings in which to train communication and collaborative skills [3]. The lack of rigor for training cultural and linguistic differences is another problem that has not been adequately addressed [3]. Both universities and companies must therefore invest in effective training to prepare students and employees for GSD challenges, specifically, those related to communication, collaboration and coordination.

In this paper we report on the evaluation of VENTURE (Virtual ENvironment for Training cUlture and language problems in global softwaRe dEvelopment) [4], a training environment that addresses the communication problems that arise in GSD by means of simulation. VENTURE provides students with real life practical examples of common interactions between development teams. This experience provides students with an insight into how best to interact over global distance with people who have different personalities, experience, skills and culture. In VENTURE, students interact with Virtual Agents that play a specific role in the Software Engineering process. Agents characterised by a specific culture and personality will interact with the student in order to simulate GSD scenarios. A Virtual Guide will correct the students and provide them with feedback in real time.

The objective of the evaluation is to assess whether VENTURE is able to provide training in GSD-specific skills, and can give the student increased confidence for effective communication with people of different cultures and languages. Moreover, from the perspective of the instructors, the goal is to assess whether the design of GSD-specific settings, is able to reflect realistic problems and scenarios that could happen in every day interactions between software engineers working in distributed teams.

The remainder of this paper is organized as follows: Section II explains VENTURE. Section III describes the research objectives of this work and Sections IV describes the research method followed. Section V explains the evaluation and Section VI summarises the results of the evaluation. Conclusions and future work are detailed in Section VII.

II. VENTURE

By using simulation, VENTURE aims to provide appropriate training, feedback and assessment by taking into account both behavioural patterns and cultural norms. The architecture of VENTURE is described in [4] and [5]. The operation of the Chat Simulator and Scenario Designer is summarized in this section. The architecture of VENTURE is described in [4] and [5]. The operation of the Chat Simulator and Scenario Designer is summarized in this section.

A. *Chat Simulator*

A simulation in VENTURE may be synchronous or asynchronous. In the case of the synchronous interactions, the student'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 student executes a Chat Simulation, a workflow engine loads the definition of the training scenario, and an introduction to the scenario is presented, after which the simulation begins. In the evaluation, a student (playing the role of a software developer) interacts with a Virtual Agent (Raúl, a Spanish software developer). A Virtual Guide (Maria) guides the student through the interaction. Maria gives immediate feedback to correct mistakes made by the user, or explain good or bad practice in more detail. Simulations are time limited, and students must therefore obtain as much information as possible from the Virtual Agent within the time limit.

B. *Scenario Designer*

Instructors can create scenarios using a Scenario Designer, which consists of a desktop application that allows the definition of the settings in the training scenarios such as: information concerning the Virtual Agents involved, the duration of the simulation, the title, and a description of the simulation.

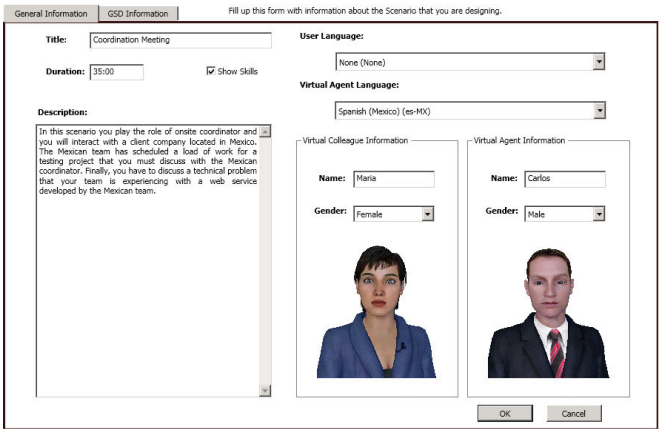

Figure 1. Scenarios Designer

Figure 1 shows a fragment of the Scenario Designer in which it is possible to define the flow of the conversation by dragging the phases from the left hand side. In order to facilitate the definition of the training scenarios, the cultural and linguistic rules can be retrieved 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 in order to be reused in the future. More details of VENTURE can be found in [4] and [5].

III. Research Objectives

The main objective of the evaluation was to obtain feedback about the environment and assess whether it is able to provide training in GSD-specific skills, and therefore, has the potential to give the student increased confidence for effective communication with people of different cultures and languages. From the perspective of the instructors, the goal was to assess whether the design of GSD-specific settings, is able to consider realistic problems and scenarios. The research questions that guided this evaluation are:

RQ1: Has the tool the potential to help participants to develop the skills needed in GSD?

RQ2: How should this tool can be applied in educational settings?

RQ3: Has the tool the potential to be effective for training purposes?

RQ4: What kinds of scenarios is this tool suited to?

The research objectives of this study are threefold: to analyse the proposal in order to implement improvements; to evaluate the adequacy of the tool to provide training in GSD, and to elicit feedback from practitioners and researchers.

IV. Research Method

Bearing in mind the research objectives described in the previous section, the research method applied is summarized in Figure 2. At the outset of this research, we conducted a systematic literature review on teaching and training in GSD [3]. At this stage we identified the teaching and training methods employed and the proposals that exist in both universities and companies. Typical problems mentioned in these studies related to the difficulties of reproducing GSD settings in educational environments, along with the difficulties of realistically reproducing specific problems in such settings.

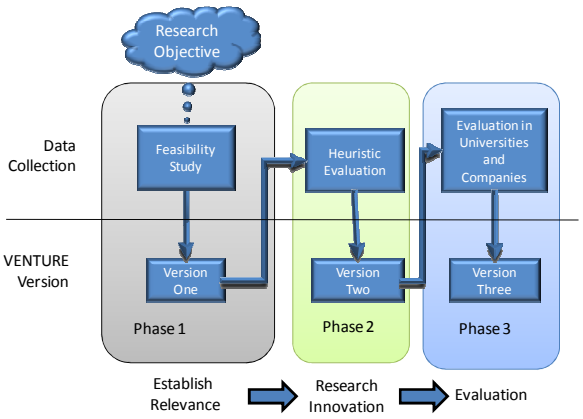

Figure 2. Research Method

After considering these results, the motivation of our research was focused on minimizing the problems found in order to provide an accurate training in specific GSD skills. The next step consisted in designing VENTURE´s architecture, which is intended to use simulation as a means to provide GSD training. An initial feasibility study was conducted in [6], in which a set of expert participants provided feedback on the Chat Simulator and the Scenarios Designer. In this paper we present an Heuristic Evaluation (Phase 2 of the research in Figure 2). Output from the Heuristic Evaluation lead to Version 2 of VENTURE.

V. EVALUATION

We have followed the Heuristic Evaluation procedure as used in [8], in which a small number of experts representing varying disciplines individually try out the tool and evaluate the interface [9]. In our case, participants are selected from complementary disciplines: research, teaching and GSD practice. An Heuristic Evaluation tests the application of recognized and accepted principles, intuitiveness, usability early in the development process, effective identification of major and minor problems, and rapidity [7].

The Heuristic Evaluation method fits with our objectives, as each expert had the opportunity to use the tool and provide feedback based on their experience. The need for this evaluation arises from the complexity of GSD, which makes it difficult to evaluate the adequacy of the training environment for providing accurate training on GSD-related skills. As a consequence, the opinion and feedback of experts is needed before launching the environment in real educational settings.

However, in our case we do not only focus on usability factors, but also on the orientation and adequacy of the tool for training in GSD-related topics. In practice therefore, we extend the Heuristic Evaluation procedure, to include questions that relate specifically to GSD training requirements (as framed by our study's four research questions).

Although there is no consensus as to the optimal number of experts required to evaluate a system [7], Hwang and Salvendy [10], indicate the ideal number to be between eight and twelve. Our study involved twelve participants from ICGSE 2012 and six from Lero.

As is shown in Table 1, the evaluation was conducted by eighteen participants from 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, and they had, on average, more than seven years of experience in GSD.

Table 1. Evaluation sample demographics

	No. Participants	Average of experience (years)	
Lero	6		4.2
ICGSE	12		10.16
Total	**18**	**Average**	**7.9**

As depicted in Figure 3, eight of the participants (44%) were classified 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.

The practitioners in our sample represented several roles, to include, project manager, general manager, researcher (in two cases), marketing manager and quality and process manager. The headquarters of these companies were located in the following four countries: India, Finland, Brazil and Germany.

Two of the seven researchers in our sample are university professors well known for their published GSD research.

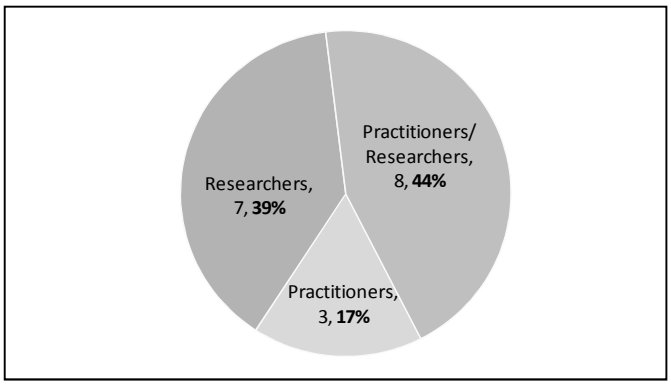

Figure 3. Profession of the participants

Table 2 describes the size of the respondents' companies, the size of their IT departments, and the number of countries usually involved in GSD activities. Out of the six companies represented three were multi-national, and two were large, with one small to medium sized enterprise. Four countries were represented through the participation of these organisations.

Table 2. Characteristics of the companies of the practitioners interviewed

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

A. *Survey procedure*

The main scope of this paper is to present the results of the Heuristic Evaluation, and its analysis will serve to implement an improved second version of the VENTURE environment. This analysis is based on the study of the responses from the Heuristic Evaluation survey (Appendix A) in addition to the study of the log of how each participant interacted with the system.

For this evaluation the Chat Simulator and the Scenario Designer were used by the participants in order to answer the research questions. A questionnaire containing open-ended and closed questions was completed by the participants. The questionnaire used in the evaluation is shown in Appendix A.

Prior to answering the questionnaire, the participants were given a verbal explanation of the objectives of the evaluation and the operation of the Chat Simulator and the Scenario Designer. They then executed a short training scenario using the Chat Simulator on their own.

The first step before starting the simulation consisted of displaying the description. In this step, we explained that during the simulation the student would play the role of the software developer and that he or she would be required to interact with a virtual Spanish developer.

During the simulation, the Virtual Agent (simulating that it was a native Spanish speaker) committed some typical Spanish mistakes when speaking English. The Virtual Guide explained these problems to the students and also corrected some of their incorrect interactions. For example, they were advised when they were too direct (or abrupt) in reporting a problem. However, specific problems relating to the student's specific culture were not taken into account in this training scenario, 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.

After finishing the simulation, the Scenario Designer was also shown to the participants, explaining how the scenarios were defined by dragging and dropping activities, and how the rules were added to the activities. They examined the definition of the scenario that they had executed, discovering its settings and the rules that they had fired. Finally, the participants completed the survey and provided their opinion and feedback.

VI. RESULTS

During the execution of the scenario, only five participants used the option to skip the current phase of the conversation. On some occasions this was owing to the fact that they attempted off topic conversations with the Virtual Agent and then they did not know how to continue the flow of the simulation.

Most of the participants (twelve) fired at least one cultural rule, which allowed them to check how the Virtual Guide corrected their inappropriate interaction. No linguistic rules were fired owing to the participants' high level of English. Table 3 also shows the participants' responses to the following questions in the questionnaire which they had to evaluate from 1 (strongly disagree) to 5 (strongly agree):

- **Q1.** The Chat Simulator is easy to use
- **Q2.** The Chat Simulator can be used to train people to recognise linguistic differences in GSD
- **Q3.** The Chat Simulator can be used to train people to recognise cultural differences in GSD
- **Q4.** The Chat Simulator is effective in correcting a student's mistakes (e.g. use of ambiguous language or inappropriate behaviour)
- **Q5.** The information in the final report provides students with useful guidance – i.e. they can, given this information, reflect on how they can improve their linguistic and cultural communication.
- **Q6.** The Designer can highlight typical cultural and linguistic mistakes in GSD
- **Q7.** I enjoyed using the Chat Simulator
- **Q8.** Using a Chat Simulator is a good way to train individuals (as a concept)

The results presented in Table 3 show that most respondents agree that the simulator is easy to use (Q1). Most of them enjoyed the experience (Q7) and believe that it is a good way to train GSD skills (Q8).

Q2, Q3 and Q4 were not given such high values, mainly because more participants interacted properly and did not fire any cultural or linguistic rules during the simulation, thus signifying that they could not be sure about the effectiveness of the tool in training these problems. Those who fired at least one rule valued these points positively. Although Q6 was also related to the previous questions, in this case the question refers to the Designer, and in this case the participants could see how the rules are added and modified.

Table 3. Results of the open questions

Participants	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
1-ICGSE	5	4	3	5	3	5	5	5
2-ICGSE	5	3	3	2	4	3	4	5
3-ICGSE	4	3	4	4	3	3	5	4
4-ICGSE	4	2	3	3	3	4	4	5
5-ICGSE	5	3	3	4	2	4	4	3
6-ICGSE	4	3	3	3	4	3	3	3
7-ICGSE	5	3	4	NA	5	4	4	3
8-ICGSE	5	4	4	3	4	3	5	4
9-ICGSE	5	5	5	4	3	5	5	5
10-ICGSE	5	3	4	4	4	4	4	4
11-ICGSE	5	4	3	3	2	5	5	4
12-ICGSE	3	4	3	3	4	4	4	4
13-Lero	4	3	3	5	5	4	5	4
14-Lero	5	3	3	4	4	4	5	5
15-Lero	4	4	2	3	4	4	4	2
16-Lero	4	4	4	3	3	3	3	4
17-Lero	5	4	4	4	4	5	5	4
18-Lero	4	4	4	4	3	5	4	3
Median	5	3.5	3	4	4	4	4.5	4

The information gathered in the final report was reported as a point for improvement (Q5) and some participants provided feedback that will be shown in the following subsections.

The results of the evaluation are summarized below, concentrating on the good points, weak points and ideas for improvement.

A. Good points

The environment was considered appropriate for training in GSD. In general, the researchers liked the idea and felt that the tool could provide independent and customized training focused on specific problems. As the tool responds in a practical manner to a real training need, they found it instructive and motivational. 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 these problems during the initial stages of the project".

In terms of student experience, both the practitioners and researchers felt that the use of the Chat Simulator is similar to any other chat application. The use of Virtual Agents is seen as a good way to simulate GSD settings and provide students with feedback. One practitioner indicated that when interacting with the Virtual Agent, students 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.

With regard to the time and resources required, the environment was also well valued, as it minimizes the instructor's workload and the time required to organize courses and seek experts, when compared with traditional methods.

With regard to the responses, the flexibility of the proposal to be tailored to a specific culture's needs was seen as one of the tool's strong points, and this made the respondents believe that it would be feasible to apply the tool in educational environments. In support of this idea, they suggested a broad set of interesting training scenarios, and its extrapolation to other types of training, not only focused on GSD.

After using the environment, some practitioners agreed that with a complete database of problems and linguistic and cultural rules, this environment would have the potential to be used by inexpert developers. In the words of one practitioner: "In the future it can be a great tool with the sufficient amount of information".

B. *Weak points*

The main weak points reported were the problem of providing a sufficient number of training scenarios to enable training in a wide variety of skills depending on the student's needs. Another problem was that the time required to create new training scenarios should be as short as possible, which would require developing a sufficient set of pre-defined rules in the databases.

Some participants found inconsistencies in the Virtual Agent's answers when they interacted in a rude manner or when they attempted off topic conversation, and suggested further work in this direction to improve the robustness of the tool before its application in educational settings. One of the participants also expected to receive answers to questions that he posed that were not considered in the scenario definition. Some participants also 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, and it would therefore be appropriate to improve the robustness in this respect.

Finally, the participants suggested that future training scenarios should be tailored and configured to specific cultural needs, since the training scenario used in the demonstration was too generic and limited in time.

C. *Ideas for improvement*

After interacting with the environment, participants suggested the following points:

- Improve the usability of the Scenario Designer by automating some of the tasks related to the management of the rule database.
- Implement automatic recognition of synonyms and abbreviations during the interactions (in real time).
- Support automatic detection of inappropriate or offensive interactions.
- Provide detailed information on specific problems the student experienced during the simulation as part of the final report.
- Implement a database of cultural and linguistic problems, in addition to the e-mail simulator.
- Create a set of training scenarios that reflect real GSD problems.

All points raised by the participants will form part of the improvements that will be implemented in the near future.

VII. LIMITATIONS

This work has some limitations as regards the construct and internal validity of the Heuristic Evaluation [11]. With regard to construct validity, it must be considered that the evaluation was limited in time, and the definition of the training scenario did not take into account the specific culture of the participant, since it was not possible to consider all the cultures involved in an international conference. Moreover, 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.

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.

VIII. CONCLUSIONS AND FUTURE WORK

By considering the answers to the closed and open questions from the survey, it can be considered that VENTURE had a positive impact on the participants. The diversity of the population participating in the evaluation, including researchers, practitioners and teachers, provided a variety of opinions focused on different aspects of the proposal, such as the user interface, the usability, effectiveness for training cultural and linguistic problems in GSD and its applicability in university classes or companies. The analysis of the results suggests that the main objectives of the tool may be fulfilled as detailed in the answers to our research questions.

The general outcome of the study is that the VENTURE platform has the potential to provide useful and meaningful scenarios with which to train practitioners in GSD. However, providing effective training for both inexperienced and experienced GSD practitioners will require a deeper and more diverse knowledge base of real problems. Populating and improving the knowledge of both the scenario database and that of cultural and linguistic rules requires not only an exhaustive literature review, but also the cooperation of experts and practitioners who could include the problems that they have actually confronted. In the future we plan to develop mechanisms to make the database available to the GSD community (both researchers and practitioners) to allow them to collaboratively improve and build this knowledge.

The next phase of this research is to conduct an evaluation of the platform in real training environments based in universities and companies. Students and inexpert

practitioners are central to this evaluation, as it is through assessing and monitoring their interaction with VENTURE that we can determine whether VENTURE meets its aims: to improve communication and cultural awareness in GSD type interactions. Moreover as part of this evaluation we plan to analyze the type of person most suited to the type of learning offered by the simulated training platform VENTURE.

ACKNOWLEDGMENTS

We would like to thank all the participants in the evaluation. This work was supported, in part, by the 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). Additional support came from ORIGIN (IDI-2010043 (1-5)) funded by CDTI and FEDER, as well as GLOBALIA (PEII11-0291-5274), Consejería de Educación y Ciencia, Junta de Comunidades de Castilla-La Mancha.

REFERENCES

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

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

[3] M. J. Monasor, 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)* Princeton, NJ, USA: IEEE Computer Society, 2010, pp. 177-186.

[4] M. J. Monasor, 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),* vol. 24, pp. 707-717, 18 Aug 2011 2011.

[5] M. J. Monasor, A. Vizcaino, and M. Piattini, "An Architecture for Creating Simulators for Training Global Software Development," in *REMIDI Workshop, International Conference on Global Software Engineering,* Helsinki, Findland, 2011, pp. 90-94.

[6] M. J. Monasor, A. Vizcaíno, M. Piattini, J. Noll, and S. Beecham, "Simulating Global Software Development processes for use in Education: A Feasibility Study," in *EuroSPI (accepted)* Dundalk, Ireland, 2013.

[7] A. Holzinger, "Usability engineering methods for software developers," *Commun. ACM,* vol. 48, pp. 71-74, 2005.

[8] S. Beecham, 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)* Porto Alegre, Brazil: IEEE Computer Society, 2012, pp. 12 - 17.

[9] J. Nielsen, "Enhancing the explanatory power of usability heuristics," in *Proceedings of the SIGCHI conference on Human factors in computing systems: celebrating interdependence* Boston, Massachusetts, United States: ACM, 1994.

[10] W. Hwang and G. Salvendy, "Number of people required for usability evaluation: the 10±2 rule," *Commun. ACM,* vol. 53, pp. 130-133.

[11] R. H. Hoyle, M. J. Harris, and C. M. Judd, *Research Methods and Social Relations*, 8 ed.: Wadsworth Publishing, 2009.

APPENDIX A: QUESTIONNAIRE USED IN THE EVALUATION

A1. Do you think this training method can be effective in the accurate training of specific cultural and linguistic differences in GSD?
A2. Do you think it would be feasible to train students/members of your university/company by applying this environment?
A3. What problems did you experience while using the Chat Simulator?
A4. Can you suggest any point for improvement or new features that you would like to be implemented in the simulator?
A5. 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?
A6. Can you suggest any point for improvement or new features that you would like to be implemented in the Designer?
A7. Please, indicate your opinion as regards the following points: 1 (strongly disagree), 5 (strongly agree): Questions Q1 to Q8 presented in Section VI
A8. Nationality
A9. How many years have you worked in GSD?
A10. Position: ☐ Practitioner ☐ Researcher ☐ Teacher of these subjects_____ ☐ Other _____
A11. May we contact you again? ☐ Yes ☐ No. Email:_____
A12. May we include your name in the list of experts of this study? ☐ Yes ☐ No

For practitioners only

A13. How many years have you worked in the Software Industry?
A14. What is your current role?
A15. Size of company?
A16. Size of IT dept?
A17. No. of countries involved in GSD, and country where head office is based?