A REVIEW OF COLLABORATIVE VIRTUAL ENVIRONMENT (CVE) EVALUATION METHODOLOGIES

Thrasyvoulos Tsiatsos[1, 2] and Konstantinidis Andreas[1]
[1]*Department of Informatics, Aristotle University of Thessaloniki*
Thessaloniki, Greece

[2]*Research Academic Computer*
Technology Institute, Greece

Keywords: Collaborative, Virtual Environment, Evaluation.

Abstract: This paper presents several methodologies concerning the evaluation of Collaborative Virtual Environments (CVEs). In doing so, the authors aim to compensate for the absence of a CVE evaluating standard, with the goal of aiding future evaluators in their task. Initially, the paper discusses the general benefits of CVEs and computer supported collaborative learning (CSCL). Based on the examined research it continues with suggested evaluation criteria and the division of evaluation into partitions. Following that, we discuss recommended evaluator profiles and data mining practices. The paper concludes with a description of appropriate data analysis procedures. In the final section, some closing remarks are made and future work is discussed.

1 INTRODUCTION

A Collaborative Virtual Environment (CVE) is a computer-based, distributed, virtual space or set of places. In such places, people can meet and interact with others, with agents, or with virtual objects. CVEs might vary in their representational richness from 3D graphical spaces, 2.5D and 2D environments, to text-based environments. Access to CVES is by no means limited to desktop devices, but might well include mobile or wearable devices, public kiosks, etc (Churchill et. Al., 2001) It is true that CVES will play an important role in future education since continuous enhancements in computer technology and the current widespread computer literacy among the public have resulted in a new generation of students that expect increasingly more from their e-learning experiences. To keep up with such expectations, e-learning systems have gone through a radical change from the initial text-based environments to more stimulating multimedia systems.

The evaluation phase in a software application's development cycle is of paramount importance. Through evaluation, software houses can cut down on development cost and time. Evaluation can also aid in the demonstration of a program's benefits to funding sources and can improve its overall effectiveness and appeal. Better evaluations can lead to better designs, based on users' and experts' recommendations and suggestions. On the other hand, as (Turner and Turner, 2002) among others have observed, the evaluation of collaborative applications is fraught with difficulty.

There reason for this difficulty will become clearer in the next section, where we discuss the multitude of criteria that a thorough and complete CVE evaluation must take into account. In section 3, we present methodologies on how to simplify the evaluation process by breaking it down into individual partitions while in the remaining sections we suggest evaluator profiles, data mining practices and data analysis procedures based on what the majority of the examined research proposes.

2 EVALUATION CRITERIA

It is important when planning an evaluation to determine which items are assessable. This is often the most complex part. This collection of items is necessary to formulate specific questionnaires and hence to find and eliminate disturbance factors from the implementation of a CVE (Goebbels et. Al., 2003)

Tsiatsos T. and Andreas K. (2008).
A REVIEW OF COLLABORATIVE VIRTUAL ENVIRONMENT (CVE) EVALUATION METHODOLOGIES.
In *Proceedings of the Third International Conference on Computer Graphics Theory and Applications*, pages 499-504
DOI: 10.5220/0001098604990504

In our research we discovered that these criteria varied considerably between evaluations. We will briefly present here the common criteria used among the majority of researchers. These include: *effectiveness, transparency, confidence, usability, interaction, application, collaborative work, system related criteria and the sense of co-presence*.

The ISO 9241-11 (ISO, 1998), defines *effectiveness* as a human's accuracy and completeness in doing tasks. A tight relation between effectiveness and efficiency is also confirmed, as efficiency is defined as the effort necessary to achieve effectiveness. A system is designated as *transparent* if the user recognizes whether the dialog system is processing an input command or is waiting for a new command.

Even though, members of the army community as well as major organizations such as Boeing, Chrysler and General Motors are now regularly using CVES for system and product development life-cycle (DLC) activities (Lethinen and Hakkarainen, 2001), there is still the matter of *confidence* in such a novel technology. In that, according to (Turner end Turner, 2002): the CVE must show that it can deliver safety-critical training to senior professionals; the training through a CVE must be validated by a recognised training and standards body as being of a suitable standard; the CVE must be accepted by the trainers, trainees and employers who will have to use it.

Usability inspections of the initial applications are necessary so as to uncover the main design flaws and allow a clean up of the design, meanwhile adapting the method to 3D collaborative aspects. Usability and interaction are very much interrelated. Concerning *interaction*, social-arbitrary knowledge (language, values, rules, morality, and symbol systems) can only be learned in interactions with others. Several human-computer interaction rules for display design must be taken into account when implementing any e-learning system, such as consistency of data display (labelling and graphic conventions), efficient information assimilation by the user, use of metaphors, minimal memory load on user, compatibility of data display with data entry, flexibility for user control of data display, presentation of information graphically where appropriate, standardized abbreviations, and presentation of digital values only where knowledge of numerical value is necessary and useful.

Application criteria are generally concerned with the affordances of objects and the lack of help with the CVE itself. They are broad in nature, from problems with objects whose operation is not obvious, to wider topics such as how best to represent group services to group members.

The difficulty in evaluating *collaborative work* is that some tasks are less "shareable" than others. For instance, solving anagrams can hardly be done collaboratively because it involves perceptual processes which are not easy to verbalise (if they are open to introspection at all). In contrast, some tasks are inherently distributed, either geographically (e.g., two radar-agents, receiving different data about the same aeroplane), functionally (e.g., the pilot and the air traffic controller) or temporally (e.g., the take-off agent and the landing-agent).

High *system* responsiveness is perceived as having very positive impact on collaboration (Goebbels et. Al., 2003). Even downsizing the application in order to decrease the CPU load is thought to be recommendable. Apparently, good system responsiveness is guaranteed if all inputs and outputs are processed and rendered within less than 50ms. Given the user's expectation of free movement at all times, a low system responsiveness suggests to the user that an error has occurred, or that the operation failed. This is also potentially serious for immersed users since the visual and proprioceptive cues will conflict.

Finally, researchers can organize isolated auxiliary case-controlled experiments focused on the evaluation of factors of the central CVE concept of *presence*. Research (Goebbels et. Al., 2003) has shown that the perception of co-presence is interrelated with the video frame rate. Further experiments with the video frame rate as a parameter showed that the perception of co-presence vanishes completely if the video frame rate sinks below 12 fps.

Additional criteria based on the conversational framework presented and discussed in (Lethinen and Hakkarainen, 2001) are resource negotiation, adaptation, monitoring, student reflection, extensibility, coordination of people and activities, individualisation and learner centeredness. Other criteria, mentioned but not shared between the researchers are input devices, physical equipment and cabling, frequency with which the user looks at the partner and frequency with which the user speaks with the partner.

Considering all these evaluation items in one session is almost impossible, since the items mentioned above evaluate too many different aspects of Human-Computer-Human interaction. In order to address this number of items special partitions of evaluation must be defined. This is the theme of the next section.

3 EVALUATION PARTITIONS

As we saw in the previous section, the broadness of CVE evaluation criteria constitutes an obstruction in the evaluation process. This has been circumvented by many researchers who propose specific criteria partitioning. These partitions differ greatly in number, purpose and scope. Some partitions simply help categorize the criteria while others constitute distinct phases or sessions of the evaluation process.

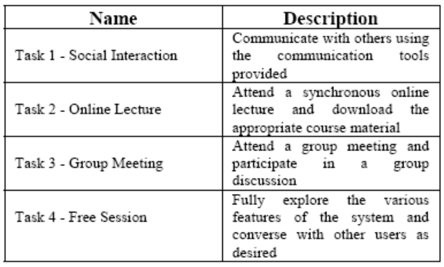
Some researchers (Michailidou and Economides, 2003) categorize the criteria as psychological, pedagogical, technical, operational, organizational, economic, social and cultural. Others such as (Turner and Turner, 2002) divide the criteria in dimensions. These dimensions are the usability dimension, the collaborative work dimension and the 'fit for purpose' dimension. The usability dimension includes many of the usual issues of interaction with the user interface (UI), such as: if users can find functions, perceive the effect of their actions and use a range of input devices. A final consideration at this level is the affordances relating to the fidelity of the virtual world to its physical counterpart, presence (meaning the sense of being in the virtual world) and engagement (meaning the sense of being 'wrapped up' in any action that may be occurring) have a close but somewhat complex association with embodiment. The collaborative work dimension is about the users trying to work through the UI to employ specific functions to collaborate with others in the environment. Finally, the 'fit for purpose' dimension is closely related with the confidence criterion presented in section 2. The researchers mentioned that users also appeared to find this partition a natural one, thus allowing meaningful discussion of pedagogic effectiveness whilst acknowledging that ergonomic issues were still outstanding. They commend the approach to others working in similar evaluation contexts.

On the other hand, some researchers such as (Grasso and Roselli, 2006) divide the criteria into three sessions. These are a usability session, a co-presence session and a co-work session. We will briefly present here their proposed methodology. In an introductory session the evaluators are informed about the display system, the equipment and the environment they are going to work with (the objective being to create almost the same conditions for all evaluators). Following that, the usability session begins. Here, the users interact autonomously within the CVE for about five minutes. During the interaction an external observer is taking notes and filling out a special observer questionnaire contributing to data mining, as will be discussed in section 5.

After this, in the co-presence session, the user works again in the CVE but now with another data set. In contrast to the latter session an experienced user who has been involved in the development process is remotely present within the same environment through an audio/video connection. The experienced user explains the task, the data set, the input devices and the tools remotely to the evaluator. Finally, in the co-work session, two users must work together to complete a task. The task is designed in such a way as to be impossible for it to be completed by just one user.

Another way of partitioning the evaluation process is by designing task-oriented social situations to represent typical collaborative scenarios. Two types of situations are defined. The first is called presentation and occurs whenever there is an expert who wants to instruct a novice, as in cases of training or in the field of education. The other is the joint work of two people who try to benefit from their combined expert knowledge in order to solve a difficult problem. A variation of this method consists of two phases: a usability inspection and a scenario based evaluation.

A final method of making sure all the criteria are assessed in the evaluation process is to expose the users to the different features of a system through specific tasks. For example, in some papers, four tasks are defined as presented in Figure 1.



Name	Description
Task 1 - Social Interaction	Communicate with others using the communication tools provided
Task 2 - Online Lecture	Attend a synchronous online lecture and download the appropriate course material
Task 3 - Group Meeting	Attend a group meeting and participate in a group discussion
Task 4 - Free Session	Fully explore the various features of the system and converse with other users as desired

Figure 1: Task oriented evaluation.

4 USER (EVALUATOR) PROFILES

The choice of evaluators depends greatly on the intended target group of the CVE and their ability to cooperate with the CVE designers. For example, when working with children it is especially important to inform them that their performance will be evaluated by the teachers as part of their class-work. Motivation is an important factor that can question the validity of experiments with children (Grasso and Roselli, 2006).

There are several other factors which must be taken into consideration when deciding upon the evaluators of a CVE. Factors such as the number of evaluators employed, previous experience with CVEs, educational background, profession and age. Previous experience is very important, since one can yield considerable detailed suggestions for improvements from CVE experts. Secondary factors include the ability to remain unaffected by virtual reality-induced symptoms and effects (VRISE), since empirical testing confirms that virtual reality systems induce physical symptoms and effects in VCES (Bochenek and Ragusa, 2003). However, this should concern mostly 3D CVE designers.

In addition, further research is needed in assessing social discomfort levels generated in CVEs, caused by participants working concurrently with real people and their avatars. Finally, when evaluating a CVE in groups it is important to consider that low achievers progressively become passive when collaborating with high achievers and that groups of three are less effective because they tend to be competitive, whilst pairs tend to be more cooperative (Dillenbourg et. al., 1996).

In the majority of the researched work, participation was anonymous and the number of evaluators varied between 10 and 30. The age of the evaluators was from a minimum 17 years to a maximum 58 years while the majority was between 22 and 27 years old. Most of the evaluators were university students, whereas there was diversity in the other users' professions.

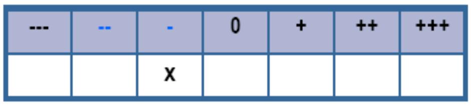
In one research (Turner and Turner, 2002) however, because of the restricted availability of users from the target community, preliminary system evaluation was carried out by 'proxy' subjects who represented the eventual user population as closely as possible in terms of relevant background skills and experience. This allowed the researchers to conserve the scarce resource of 'real' users for more polished versions of the software. In the same research it is argued that at least where issues of trust and confidence are involved, domain-specific techniques can only be developed with the participation of the community concerned.

Selecting evaluators more befittingly can lead to the more efficient acquisition of data during the data mining process. Accuracy in the acquired data, translates to a more conclusive and convincing analysis. These subjects will be discussed in the next sections.

5 DATA MINING

This section discusses the tools and methods designers can use to gain important information from the evaluators' experience with the CVE. Separating the evaluation procedure into subjective and objective is suitable when comparing two different conditions like a non-computerized approach (e.g. the traditional textbook) and a contemporary one (e.g. CVES). The objective part takes into account how the subjects interact and deal with the different features; and can be recorded automatically by the software. The subjective part reflects the user's impression about the interfaces, and can be acquired through surveys and interviews.

Data collection usually consists of three stages: pre-testing, system interaction and post-testing. Most common among researches (e.g. (Turner and Turner, 2002), (Dillenbourg et. al., 1996)), is the filling out of questionnaires or web-based forms before and after the evaluation process by the users. The pre-questionnaire is usually about self-rating or demographic information, while the post-questionnaire concerns itself with the overall CVE experience and criteria mentioned in section 2. It should be noted that if the questionnaires are similar in nature, in order to avoid carry-over effects the wording of the post-test should be slightly different. Within the questionnaires, multiple choice questions, free answers and Likert scales (a variation of which can be seen in Figure 2) are regularly used.



---	--	-	0	+	++	+++
		x				

Figure 2: A variation of a Likert scale called a seven-point scale.

The multiple choice questions are generally grouped into sections such as: general impression of the collaborative system; details about virtual worlds and avatars; navigation within the system; cooperation modes and interaction with the collaborator; technical questions about the interaction elements.

On the other hand, the free questions are concerned with general ideas for improved usability of the cooperation modes and omitted information the participants felt they required about each other in order to solve their tasks faster. After the questionnaires are examined, video recorded interviews can be carried out to clarify certain points.

Specifically in (Turner and Turner, 2002), the researchers based their set of tasks and questionnaire items on Laurillard's model of teaching and learning (Laurillard, 1993) The final version of their software was evaluated by experienced tutors. The NASA ITQ questionnaire (a measure of immersive tendencies) was administered before the evaluation started followed up by a questionnaire instrument incorporating the collaborative and pedagogic aspects, coupled with the NASA PQ – the counterpart to the ITQ which aims to measure presence.

As mentioned in section 3, when users interact with the CVE, they can be videotaped as observers monitor their progress, supported by checklists mirroring the questionnaire content. These can be used for post-evaluation discussions derived from analysis of the verbalisations and behaviour recorded.

The examined evaluation procedures took place on an average of three consecutive days and had a mean total evaluation time of three hours.

6 DATA ANALYSIS

The analysis of the acquired data will establish the data mining process' validity and lead to conclusions which will help advance CVE development and design. In order to deal with values between zero and one, the acquired data should be normalized for simplifying further operations of statistical analysis by SPSS, Origin or any other software program used.

The t-test technique is suitable for user numbers less than thirty. In this test, given two data sets, each characterized by its mean, standard deviation and number of data points; we can determine whether the means are distinct, provided that the underlying distributions can be assumed to be normal.

While the t-test is suitable for independent samples, the one way analysis of variance (ANOVA) fits for tracing a number of groups on one variable, like finding out the effect of students' learning depending on different studying methods (King et. al., 2003) ANOVA is a collection of statistical models, and their associated procedures, in which the observed variance is partitioned into components due to different explanatory variables.

Another proposed method of analysis is presented in (Laurillard, 1993). Here the analysis is partitioned into three levels. In the first level analysis, average values and their expectancy values are computed and compared for each session separately. Following that, in the group analysis, these statistical values are compared between the different sessions. Since the questionnaires were especially designed so that questions belonging to different sessions evaluate similar criteria. Finally, in the variation group analysis there is again a comparison of different sessions with each other. The difference here lays in the alteration of specific factors between groups in order to cross-check the influence of these particular factors in supporting team work.

Finally, some researchers analyze the collaboration into two levels. One level covered the technical aspects and issues the participants were faced with, while the other dealt with social aspects of the collaboration, like how the participants interacted with each other. For this analysis methods from psychological discourse analysis and sequential film analysis were used.

7 CONCLUSIONS

In this paper, based on the methodology the majority of the examined research is adopting, we presented the most important factors any CVE designer must take into consideration when planning an evaluation procedure. Our goal was to aid future evaluators in coordinating more efficient and conclusive research. We described in brief the evaluation criteria engaging the majority of examined researchers and their propositions in organizing these criteria into processable modules. Following that, we discussed the profiles of the users (evaluators) and how the correct selection of these users can lead to more accurate data mining and more precise data analysis. In the last sections we presented specific data mining tools and analysis procedures incorporated by the majority of the researched work. Our next step is to keep examining CVE evaluation methodologies with the prospect of suggesting an effective evaluation practice.

REFERENCES

Bochenek, G.M.; Ragusa, J.M.System Sciences, Virtual (3D) collaborative environments: an improved environment for integrated product team interaction?, Proceedings of the 36th Annual Hawaii International Conference on Volume, 6-9 Jan. Page(s): 10 pp (2003)

Britain, S., and Liber, O., A Framework for the Pedagogical Evaluation of eLearning Environments, (2004), http://tinyurl.com/2wdf3x

Churchill E., Snowdon D. and Munro A., Collaborative Virtual Environments: Digital Places and Spaces for Interaction, Springer-Verlag, London Limited, Great Britain, (2001).

Dillenbourg, P., Baker, M., Blaye, A. and O'Malley, C.The evolution of research on collaborative learning. In E. Spada & P. Reiman (Eds) Learning in Humans and Machine: Towards an interdisciplinary learning science. (Pp. 189- 211). Oxford: Elsevier, (1996).

Goebbels, G., Lalioti, V., Göbel, M., Design and Evaluation of Team Work in , Distributed Collaborative Virtual Environments, Virtual Reality Software and Technology Proceedings of the ACM symposium on Virtual reality software and technology, Osaka, Japan , p. 231-238, ACM Press, New York, USA, (2003).

Grasso, A., and Roselli, T., Cooperative Student Assessment Method: an Evaluation Study, International Journal of Emerging Technologies in Learning (iJET), Vol. 1, No. 2, (2006).

ISO 9241-11: Guidance on Usability, (1998).

King, Bruce M., Minium, Edward W., Statistical Reasoning in Psychology and Education, Fourth Edition. Hoboken, New Jersey: John Wiley & Sons, Inc. ISBN 0-471-21187-7. (2003)

Laurillard, D. M., Rethinking University Teaching: A Framework for the Effective Use of Educational Technology. Routledge, London, (1993).

Lehtinen, E. and Hakkarainen K., Computer Supported Collaborative Learning: A Review, (2001), http://tinyurl.com/226965

Michailidou, A., and Economides, A., E-learn: Towards a Collaborative Educational Virtual Environment, Journal of Information Technology Education, Vol. 2, pp. 131-152, (2003).

Turner, P. and Turner, S. An affordance-based framework for CVE evaluation. To appear in People and Computers XVI, Proceedings of HCI'02, (2002)