Abstract. Computer-assisted instruction has been around for decades. There has been much speculation about the benefits of computer-mediated learning. Numerous applications have been developed in different domains incorporated with emerging technologies. In recent years, advanced technologies, such as Augmented Reality (AR) and Virtual Reality (VR), have received much attention in their potential of creating interactive learning experience for the users. However, related literature and empirical studies indicated that learning effects in computer-simulated environments or Virtual Environments (VEs) are not systematically tested. Furthermore, the performance and learning in computer-simulated learning environment need to be evaluated through more rigorous methods. This paper suggests that 1) the efficacy of VEs is subject to a close examination, not only in terms of how VE-based training systems are easy of use, but also in terms of how effective learning is; 2) evaluation of learning in computer-simulated learning environments is required to be reconsidered in terms of theoretical basis and evaluation methodologies that are relevant to the measurement of training effectiveness in computer-simulated virtual learning environment. This paper explains on how learning can be assessed in VEs through the lens of training evaluation.

1. INTRODUCTION

Computer-simulated environment or Virtual Environment (VE, also known as virtual reality, cyberspace) is a computer-generated, 3D spatial environment that can help people in understanding or learning complex concepts or tasks in a safe environment, which can be fully immersive, non-immersive or augmented [12].

VR and AR as emerging simulation technologies have been developed in different domains for training, education, entertainment, marketing, collaboration, and community building purposes [8, 12].

Many [8, 21, 23, 32] claimed that the benefits and potentials of interactive and immersive quality of VR and AR technologies for education and training are numerous. Dalgarno et al [9] support this view and state that 3D virtual learning environments have a unique set of characteristics (immersion, presence, fidelity and learner control), which have potential to offer superior learning experiences. Hence, learning of both complex conceptual knowledge and motor skills can be achieved effectively in 3D VE. In their study, Dalgarno et al [9 p.152] provide a list of contributions of 3D VE for learning, for example, “facilitate familiarisation of inaccessible environment; facilitate task mastery through practice of dangerous or expensive tasks; improve transfer by situating learning in a realistic context, improve motivation” etc. These contributions emphasise the benefits and potentials of VE for learning and training, which have also been reported elsewhere [3, 21].

Previous research indicates that to design effective VE and exploit its full potential for training and education, evaluation of the efficacy of VEs are necessary [34].

This review will cover three main sections. First section addresses some VR applications in training and education, current design and evaluation challenges will also be addressed. The second section discusses various methods for evaluation of VEs for training and education. Limitations and strength of different methods will be discussed. The third section explains an evaluation matrix constructed for the purpose of determining the efficacy of VEs. Underlying theories and models for constructing this matrix will be described and discussed, follows by concludes the paper.

2. THE USE OF VEs FOR TRAINING AND EDUCATION

2.1 VR Applications

One of the most popular uses of VR is in the context of training [18]. Many VR applications have been designed, developed and implemented for training and education purposes.

Various domains have integrated VR-based training systems for its professionals. For example, in medical field, many VR simulators have been developed for training of surgical skills [29]; in manufacturing settings, VR-based training systems are established for training of object assembly operations [32, 28, 6]; and in aerospace industry, VR-based simulators have long history of been used in preflight training [1].

VR applications have also been designed for different age groups, different purposes, and tasks in training and education apart from aforementioned fields. For example, Bowman et al [2] have developed a virtual zoo exhibit (based on a previous VR gorilla exhibit application) to teach middle school students about gorilla behaviors, vocalizations and interactions in the virtual habitat. Loftin et al [17] used VEs to simulate
military checkpoint duties and involved participants in becoming part of those simulations. Finally, Dunser et al [10] developed an interactive VE, Construct3D, as a spatial ability training tool used by high school (average 17 years old) students.

The popularity of using VEs for training and education is obvious. One reason contribute to this trend may due to the antecedents and effects of immersion that VEs support. Rose [25] suggests that when compare with traditional instructional media, VR or AR technologies promote more learner attention, engagement, and motivation, which could be highly helpful to promote skills and knowledge of learner. Another reason contribute to the grows of VE-based training reported by Chen [7] is that, VR and AR technologies enable learner to interact with his or her learning environment with their hands and body, “sensorimotor feedback” during user-interface interaction may enhance learning experience [7].

It is neither to say that VEs are at by all means better than traditional instructional media, nor VEs are suits for all types of training and educational purposes. In line with others [34, 27], this paper argues that to obtain an objective view of appropriateness of VEs for training and education, evaluation of efficacy of VEs is both necessary and in urgent need. Yet, challenge of establish an effective and valid measure of VEs efficacy is what researchers, designers, educators are currently facing.

2.2 Challenge of Design And Evaluation Of VEs

With the increasing demands of using VE, designers are facing a big challenge in achieving effective VE systems design. A lack of understanding of human-computer interaction in VE [26], especially, human-VE interaction (HVEI), did not help designers to ameliorate the situation. This leads to further problems of what actually contribute to learning in technology-mediated learning context? and how to better design technology-mediated learning environment to facilitate and enhance learning? [10]. Consequently, questions of how VE-based training and learning can be made more effective and efficient need to be addressed.

Evaluation of the efficacy of VEs may be a solution, but it is difficulty task. In particular, due to the lack of empirically derived validations of interaction and learning in VE systems, training effectiveness is hardly ever systematically tested and understood in VE-based training contexts [26].

Typically, researchers, educators and even designers often have an intuitive sense that interactive and immersive qualities of VE could be highly useful to promote skills and knowledge [25]. Efficacy of VEs is based on the flawed assumption that “impressive technology” leads to useful and usable systems [25, 31]. As a consequence, limited knowledge and common grounds shared among researchers, developers, educators and other stakeholders on how effective VEs are for training and learning.

Assessment in training is a form of quality assurance in the future and is inseparable from the broad goals of training [20]. However, a few studies have been devoted to the evaluation of training effectiveness; learning is poorly defined, understood, and assessed during training process [15].

The following sections concerning the remaining questions in the field of simulation training: how effective the simulation technologies are for training and learning? What theories and methods are suitable to assess learning in VR simulation training? What are the effective evaluation methods applicable to measure VE efficacy?

3. ASSESS THE EFFICACY OF VR SYSTEMS FOR TRAINING AND EDUCATION

3.1 Studies and Methods

Rose [25] attempted to develop a paradigm to guide evaluation of VE for learning. Instructional factors, virtual environment experience factors and external factors are identified that are relevant to measure effectiveness of virtual learning environments. This research project indicates that through identifying important factors that influence learning effectiveness, a comprehensive assessment of efficacy of VE may be developed. Another early attempt to evaluate learning experience and achievement in VEs was made by [19]. Empirical based comparative studies of learning in traditional classroom setting and learning through VR are conducted. Surprisingly, learning results appear to be better in traditional lecture based instruction than learning through VR, yet the difference is not significant. This research leads to more questions than its findings in regards issues of learner preferences, enjoyment and outcomes in VE context. Recently, Chittaro & Ranon [8] proposed a different approach for evaluating VEs. Three factors are considered important measures of VE efficacy: understanding, transfer of training and retention.

The practical significance of evaluating efficacy of simulation technologies are recognized in a number of fields of research, particularly in medical domain, where use of VR simulators for training and education have increased dramatically and growing. Various methods for conducting efficacy evaluation of VR simulators are proposed by researchers related to this field. For example, Richard [26] proposed an objective assessment metrics to assess the efficacy of VR simulators for training technical skills. Sherman et al [28] focused on development of summary metrics for VR simulator and assess the construct validity. Srinivasan et al [30] conducted a quantitative based meta-analysis to assess training effectiveness of VR simulators. Of these studies, time on tasks, error rate, and completion rate of tasks are also essential measures for evaluating VE efficacy. Quantitative methodological approaches are in favor among these studies.
Even though these researches shed some light on “how” to conduct evaluation of VEs efficacy, lack of details in baseline measures made them impractical for conducting evaluating the efficacy of VEs.

One suggestion [19, p.462] made is that “there is a need for an instrument that can be used to evaluate the usability of VE” so that the effectiveness of VE system can be ascertained. Such “advances would help greatly in the evaluation of using VE for learning”. Some [3, 31] usability-focused evaluation are conducted to gather users’ perspective on effectiveness of VE systems. Theng et al [32] also conducted usability focused study to evaluate effectiveness, usefulness and acceptance of VEs from primary school children point of view.

Usability focus on how a computer system (or an artefact) is easy of use, easy to learn and useful to assist people for their work/tasks. Purpose of usability evaluation, in general, is to identify usability problems of a system and use such knowledge gained to improve the system design. Based on a review and synthesis of information provided by [4, 23], this research refers usability at four dimensions: effectiveness, efficiency and satisfaction and enjoyment.

Figure 1. Definition of VE usability

Coupling usability with VEs is essential [3]. Many usability criteria have been identified [3, 31] which serve as valuable knowledge base for indentifying influential factors for evaluating VE-based training systems.

3.2 Evaluation Parameters Of VE Efficacy

By arguing that it is urgent to evaluate the efficacy of VE for training, this paper suggests that evaluation of VE should not limit to the usability aspect, rather a broader concept of ‘efficacy’ should be coupled with VE evaluation, which is more appropriate to justify learning that VE supports.

Three parameters constitute of efficacy evaluation, apart from usability, another two elements that reference to VE efficacy evaluation are learnability and feasibility. Learnability refers to how easy VE system facilitate user to accomplish required tasks or achieve learning goals during training process. It concerns users’ expectations, intuitiveness of VE system and understandably of users about VE system [23]. Feasibility refers to appropriateness of VEs for training. It concerns the quality of VEs in imitating real life experience, learning task representation and learning content construction.

4. EVALUATION MATRIX FOR VEs EFFICACY

Another problem with previous research on VEs efficacy evaluation is that there are limited theoretically based models available for training evaluation [36]. In VEs training evaluation context, such theoretical models did not exist. What learning meant, and how training evaluation relates to learning outcomes, and what are appropriate measures that applicable specific to VEs are some of the questions our research project aim to answer.

Kariger et al [15]’s “cognitive, skill-based and affective theory of learning outcomes” design for training evaluation, which we believe could be useful to guide the development of a multidimensional matrix, aims at evaluating VE efficacy in training.

According to Kariger et al [15], learning is multidimensional and may be evident from changes in cognitive, skill or affective capacities. Evaluation training effectiveness, in the context of VE, the efficacy of VEs for training, assessing learning at this three dimensions are essential. They further explained that cognitive perspective focuses on the dynamic processes of knowledge acquisition, organization and application; skill-based perspective concerns the development of technical or motor skills of learner/trainee. Affective perspective concerns learner/trainee attitudinal and motivational impact.

In the context of VE-based training, changes of learner in cognitive, skill or affective capacity are derived during user-VE-interaction or as a result of such interaction.

Factors influence user-VE-iteration would lead to impact on changes of cognitive, skill-based and affective outcomes. For example, system factors, user factors and task factors are identified the major influences on learning-skill, affective or cognitive outcomes perspectives [3]. According to [27], system factors consider VEs system interface and VE user interface and how effective they are for user to carry out tasks (e.g. learning domain knowledge or skills). User factors consider human factors and anthropology (e.g. users’ learning style, prior experience or attitudes towards computer technology). Tasks factors relate to the nature of the tasks and relate to VEs system features such as technical quality, functional and non-functional requirement of software system. Effective VEs system design that balance these factors are essential for achieve effective learning during training. Evaluation need to design appropriate measures of these factors.
Previous studies [22, 4, 27, 23] shown that usability, learnability and feasibility attributes tightly coupled with the quality and levels of human-VE interaction (HVEI), which may contributed to changes in learners in terms of cognition (information processing) capability, skill/performance level, and affective/motivation. In addition, users’ psychological involvement (in terms of perception, attention, comprehension and cognition) during HVEI formulate users’ perspective on usability, learnability or feasibility of VEs and their learning experience.

Figure 2 illustrates the conceptual model of our matrix. Apart from commonly employed objective measure of the effectiveness of VE (e.g. completion time, error rate), user-centered evaluation in favor of gathering subjective data of users’ perception of the effectiveness of VEs, their perceived quality of learning and performance as well as their interaction and experience in VEs. A compressive set of measures are developed based on the matrix, both subjective and objective data will be collected.

<table>
<thead>
<tr>
<th>Evaluation parameter</th>
<th>Psychological &amp; Cognitive process</th>
<th>HVEI impact</th>
<th>Learning outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Usability</td>
<td>Acceptance</td>
<td>Efficiency</td>
<td>Interactivity</td>
</tr>
<tr>
<td>- Effectiveness</td>
<td>Perception</td>
<td>- Effectiveness</td>
<td>- Interactivity</td>
</tr>
<tr>
<td>- Efficiency</td>
<td>- Perceived</td>
<td>- Efficiency</td>
<td>- Interactivity</td>
</tr>
<tr>
<td>- Enjoyment</td>
<td>- Perceived</td>
<td>- Enjoyment</td>
<td>- Interactivity</td>
</tr>
<tr>
<td>- Satisfaction</td>
<td>- Perceived</td>
<td>- Satisfaction</td>
<td>- Interactivity</td>
</tr>
<tr>
<td>Learnability</td>
<td>Automation</td>
<td>Engagement</td>
<td>Affective, cognitive</td>
</tr>
<tr>
<td>- Intention</td>
<td>Automation</td>
<td>Engagement</td>
<td>Affective, cognitive</td>
</tr>
<tr>
<td>- Ease of learning</td>
<td>Automation</td>
<td>Engagement</td>
<td>Affective, cognitive</td>
</tr>
<tr>
<td>- Ease of use</td>
<td>Automation</td>
<td>Engagement</td>
<td>Affective, cognitive</td>
</tr>
<tr>
<td>Feasibility</td>
<td>Comprehension &amp; cognition</td>
<td>Interaction &amp; presence</td>
<td>Affective, skill &amp; cognitive</td>
</tr>
<tr>
<td>- Experience representation</td>
<td>Comprehension &amp; cognition</td>
<td>Interaction &amp; presence</td>
<td>Affective, skill &amp; cognitive</td>
</tr>
<tr>
<td>- Task representation</td>
<td>Comprehension &amp; cognition</td>
<td>Interaction &amp; presence</td>
<td>Affective, skill &amp; cognitive</td>
</tr>
<tr>
<td>- Content representation</td>
<td>Comprehension &amp; cognition</td>
<td>Interaction &amp; presence</td>
<td>Affective, skill &amp; cognitive</td>
</tr>
</tbody>
</table>

Figure 2: Model of efficacy evaluation for VEs

5. CONCLUSION

Much attention on VEs may be due to the benefits and potentials of both technological and psychological capabilities in creating engagement, interactive and immersive learning experiences for the users.

It is obvious that more intuitive interaction metaphor can be afforded by VEs compared to traditional technologies for the user to interact with the world through multi sensory modalities.

Development of a practical assessment tool, (Figure 2) would be very helpful, not only in the design and development of effective VR-based training systems, but also can be used as a reference to predict learning outcome assessment by designers, evaluators, or even learner/trainee themselves.

Currently ISR team at Deakin University is in the process of improvement of a virtual and augmented reality system for training and maintenance in aircraft/automotive industry. Integration and validation of this matrix is underway. Empirical data will soon be available to provide initial results for validation purpose.

REFERENCES


