

# Learners' Assessment and Evaluation in Serious Games: Approaches and Techniques Review

Ibtissem Daoudi<sup>1,2(✉)</sup>, Erwan Tranvouez<sup>2(✉)</sup>, Raoudha Chebil<sup>1(✉)</sup>,  
Bernard Espinasse<sup>2</sup>, and Wided Lejouad Chaari<sup>1</sup>

<sup>1</sup> ENSI, COSMOS, Manouba University, 2010 Manouba, Tunisie  
{ibtissem.daoudi, raoudha.chebil,  
wided.chaari}@ensi-uma.tn

<sup>2</sup> Aix Marseille University, CNRS, LSIS UMR 7296, 13397 Marseille, France  
{erwan.tranvouez, bernard.espinasse}@lsis.org

**Abstract.** Recently, there has been growing interest in the use of Serious Games (SG), as they provide a more powerful means of knowledge transfer in almost every application domain especially in the crisis management field. With this increasing adoption of SG, designing novel techniques for learners' assessment and evaluation has become of paramount importance to improve learning results and thus to maintain players' motivation. This paper focuses on the learners' assessment and evaluation in SG. After defining assessment and evaluation, we distinguish two main approaches: implicit and explicit. For each of these approaches, we present some techniques currently used in some existing games. Then we compare these different approaches and techniques. This synthesis is expected to help researchers and games creators working in this area and identifying benefits and limitations of these techniques in order to develop a new comprehensive technique that outperforms all existing ones.

**Keywords:** Serious Games · Crisis management · Game evaluation · Learners' assessment · State of the art · Taxonomy of techniques

## 1 Introduction

The rapid development in Information and Communication Technologies has introduced the concept of games designed for a serious purpose other than pure entertainment so-called Serious Games (SG) [1]. The goal of the SG is to make the knowledge and/or competencies acquisition more efficient and attractive than classical learning methods. The growing interest for SG environments, especially for training, has raised new needs in terms of learners' assessment and evaluation [2]. This topic constitutes an important component of any adaptive SG as it maintains relevant information about what went right or wrong during a game session. This information is useful since it is exploited in order to provide to learners the most suitable adaptation according to their profiles and learning objectives/needs.

Crisis management represents a fertile playground for SG because of its *availability* and relative *low cost* (compared to field exercises) and the *variety of situations*

(industrial accident, forest fires, floods, terrorist attacks...) each involving *multiple roles* (first responders, chain of command, civilian officials...) and *collaborative behaviors* (evacuation, victim salvation, decision process...) [1]. This complexity offers Research & Development opportunities characterized by pluridisciplinary and inter-disciplinary contributions.

SG can range from relatively simple (linear scenario) one-shot development<sup>1</sup> supporting an information campaign (marketing oriented) targeting general public awareness to complex training framework with multi-actors scenario reproducing real crisis management situation for professional (simulation oriented)<sup>2</sup>. Somehow correlated, Crisis Management SG either can be an *ad hoc* software solution to a particular need or developed (generated) with dedicated software development environments [9]. So-called *SG generator* (Game Engine only or domain dedicated Computer Aided Software Environment), include however implicit conceptual limitations on game and learning characteristics (scenario complexity, number of players...) depending on their “target” (3D environment, web game...). Moreover, one “cultural traits” of Crisis Management is the importance of assessing post-crisis what happened, which behaviors were adequate and what went wrong in order to improve procedures and/or training. Such debriefing is also required in virtual training environment and completed by automated assessment. This assessment is more complex in a multi-actors context, multi-skills, and emotion management while keeping the players engaged in the crisis scenario. Above providing assessment capabilities, Crisis Management SGs also require to be evaluated in regards to their training capabilities.

This paper presents a survey of this research issue, describing the main techniques and proposing a taxonomy to better organize them. Section 2 defines differences between assessment and evaluation, distinguishes two approaches of learners’ assessment and evaluation, and presents a review of the main techniques used in existing SG related to explicit and implicit approaches. Section 3 compares several SG that have been assessed/ evaluated. Finally, conclusions are drawn and directions for future work are presented.

## 2 Learners’ Assessment and Evaluation Approaches in SG

### 2.1 Assessment Versus Evaluation

Both assessment and evaluation require (qualitative and/or quantitative) data about learners and utilize (direct and/or indirect) measures to understand and analyze learners’ behaviors during a learning session. However, *assessment* is defined as a process of collecting and interpreting data about learners in order to provide them feedbacks on their failures and progress and to make then improvements of their current performances; whereas *evaluation* is the process of making judgments about learners’ performances or SG effectiveness based on defined criteria [6].

---

<sup>1</sup> Such as « Mr Travel » on traveler behavior.

<sup>2</sup> Such as CRISE solutions (<http://www.vr-crisis.com>) .

For more clarity, *assessment* can be described as a “formative” measurement implemented and present throughout the entire learning process for the purpose of diagnosing learners’ actions and identifying areas of improvement to increase learning quality. *Evaluation* is a “summative” assessment conducted at the end of a learning process in the purpose to test the overall learners’ achievements and to draw judgments about learning quality. So, we can conclude that assessment is concerned with learning process, while evaluation focuses on the product (SG). Figure 1 summarizes the key differences and similarities between assessment and evaluation.

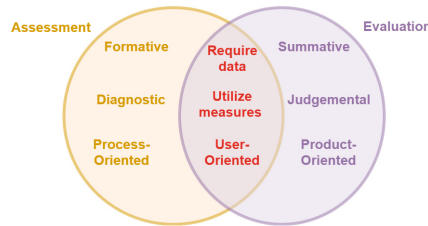


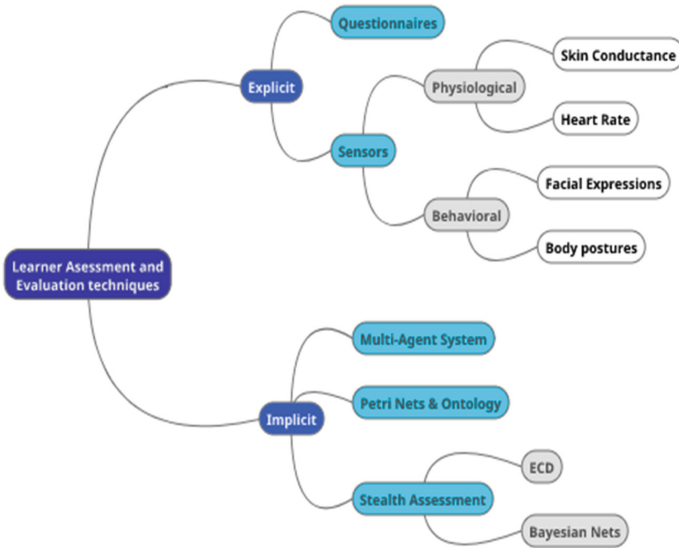
Fig. 1. Assessment Vs Evaluation (adapted from [14])

## 2.2 Taxonomy of Assessment and Evaluation Techniques

The state of the art of learners’ assessment and evaluation in SG is quite rich [6]. In this review, we propose to classify recent existing works into two main approaches according to the technique type used in assessment or evaluation process. The first approach gathers all techniques that assess/evaluate learners explicitly like questionnaires [3, 7, 13] and physiological sensors [10]. The second approach focuses on techniques that assess/evaluate learners implicitly using models and methods of Artificial Intelligence (AI) such as Petri nets and ontology [11] as well as agent technology [2]. The main difference between explicit and implicit techniques relates to the ways of collecting and analyzing data about learners. On the one hand, an *explicit approach* aims to use a direct and obvious measure of collecting and analyzing data about learners. On the other hand, an *implicit approach* aims to collect and to analyze data about learners in an indirect and unobtrusive way, without disrupting the high level of engagement provided by SG. It can be assimilated to stealth assessment [8].

To make the scope of the review more clear, we propose a taxonomy of learners’ assessment and evaluation techniques in SG. This taxonomy will structure and guide the survey of SG in the following sections. Figure 2 presents the organization of the key aspects of our taxonomy from the most general to the most specific.

**Explicit assessment** can be accomplished by using a questionnaire or some sensor devices. In fact, self-report questionnaires [3, 7, 13] are frequently employed because it is simple to implement, but it represents a subjective assessment which relies on non-exhaustive players opinions [6]. Also, questionnaires disrupt the high level of



**Fig. 2.** Taxonomy of learners' assessment and evaluation techniques

engagement provided by SG since they require stopping the learner from playing and requesting her/him to answer questions. Furthermore, the use of hardware and software equipments provides an explicit way to assess/evaluate the learner while using SG [10]. This technique can provide additional information for learner assessment in real-time without stopping him/her during playing. However, it obviously requires the use of additional sophisticated devices that can be expensive. In addition, data collected using these equipments can be interpreted in different ways, which can affect negatively the reliability of the learner assessments results [6].

*Implicit* learners' assessment and evaluation exploits the AI techniques in order to assess/evaluate the behavior of learners such as multi-agent architecture [2], Petri Nets combined with ontology [11] and the conceptual framework Evidence-Centered Design (ECD) combined with Bayesian Nets [8]. All these approaches have the major advantage of adopting implicit models and methods of AI for learners' assessment and evaluation without endangering the high level of engagement provided by SG. Therefore, this type of assessment is intended to support learning and increase learners' motivation. In a Crisis Management context, this may mean "believability" and improves the learning of procedures and best practices. However, most of these approaches consider only one criterion to assess/evaluate learners' reactions to SG adoption in a particular training process.

### 2.3 Explicit Techniques of Learners' Assessment and Evaluation in SG

Learners' assessment can be performed through evaluating the learners' answers to a questionnaire at the beginning, during or at the end of a game session [3, 7, 13]. For example, Silva et al. [13] invited children, residents of the city of Rio de Janeiro in

Brazil, to use the SG “Stop Disasters” to build a safety culture for emergencies. To assess learners’ performances and to verify if the game really improves the awareness of risky situations, the participants answered questionnaires before and after playing the SG about three main aspects namely gameplay, missions and game scenarios.

Advances in neurosciences branch have led to the development of various equipments able to detect and recognize human emotions via facial expressions and physiological signals. Several works have shown that these measures can provide an indication of learners’ emotions [10]. For instance, Mora et al. [10] showed the usefulness of collecting data from the WATCHiT sensor during a training event to support debriefing in the crisis management field by addressing two different scenarios. This debriefing, based on sensor data, is considered as a form of evaluation with explicit attention to emotions as well as ideas and behaviors of learners.

#### 2.4 Implicit Techniques of Learners’ Assessment and Evaluation in SG

Learners’ assessment while playing a serious game can be supported by the agent technology. For example, Oulhaci et al. [2] presented a multi-criteria and distributed assessment approach of learners in “SIMFOR” SG. They propose a methodological framework for learners’ assessment based on the concept of Evaluation Space allowing the production of individual and collective multicriteria assessments. In order to implement this methodological assessment framework, they have developed an agent-based architecture improving Non-Player Character (NPC) adaptability (simulation of NPC behavior) and supporting individual and collective learners’ assessment. Moreover, Shute [8] proposed an assessment approach embedded within a SG based on the conceptual framework Evidence-Centered Design (ECD) and Bayesian networks in order to model and assess important competencies. In addition, Pradeepa et al. [11] developed an assessment approach that combines a Petri Network and ontology to track not only the player’s actions but also to analyze and diagnose the knowledge acquisition of the learner.

### 3 Comparative Study of SG Assessment and Evaluation

This section describes crisis management SGs providing learner assessment/evaluation during game play. These SG are classified according to the proposed taxonomy. As shown in Table 1, the process of learners’ assessment and evaluation in SG requires several inputs collected from the learners’ interaction with the game. The techniques described in this article exploit these inputs to extract useful information about the learner(s) and to assess/evaluate his (their) behaviors (outputs).

Table 1 shows that most of SG have been evaluated using explicit techniques. For example, Stop Disasters [13], GDACS mobile [7] and DREAD-ED [3] were evaluated via learners’ answers to questionnaires. Table 1 also indicates that only the SG “SIMFOR” [2] used a multi-agent architecture as an implicit technique in order to produce individual and collective assessments. To sum up, we conclude that there is a lack of works considering the emotion concept in learners’ assessment in the context of crisis management SG. In fact, human emotions play a huge role in the process of group

**Table 1.** Serious games and learners' assessment and evaluation techniques

SG	Multi-players	Approach	Inputs	Outputs	Techniques
SimFor [2]	Yes	Implicit Assessment & Evaluation	Learner' Model	Learners' Skills & Social Interactions	Multi-Agent System
DREAD-ED [3]	Yes	Explicit Evaluation	Interview & questionnaire responses	Learners' Social Skills	Interview & Questionnaire
Stop Disasters [13]	No	Explicit Evaluation	Questionnaire responses	Learners' Skills	Questionnaire
SPRITE [5]	No	Implicit Evaluation	Player score	Learners' Skills	Scoring
GDACS mobile [7]	Yes	Explicit Evaluation	Questionnaire responses	Learners' interactions	Debriefing & questionnaire

decision making. In crisis management filed, feeling negative emotions like stress and fear has a negative impact on the individual performance of player during a crisis response. This consequence can affect negatively the collective performance of the group and thus the success of a game session. Additionally, the works exploiting the social interactions aspect in a collaborative context of crisis management games are limited [3, 4]. However, it is important to address the role of social relationships between the different actors in affecting group decision making. In fact, in order to make successful a teamed crisis management, each member should contribute equally and communicate all relevant information to others before making a joint decision.

To tackle this problem, a new technological phenomenon, called "Educational Data Mining" (EDM), proposes to explore big data capabilities in an educational context. It is defined as an emerging discipline concerned with developing, researching and applying computerized methods for exploring data that come from the educational setting and using those methods to better understand learners' behaviors and the settings which they learn in [12]. EDM can be useful in the field of crisis management SG since it manipulates heterogeneous data representing actors' actions, attitudes and interactions relating to a crisis as well as their consequences once the scenario played.

## 4 Conclusion

This paper presents an overview of Crisis Management SG focused on their learners' assessment and evaluation capabilities. This synthesis can help researchers and game creators by enlighten the main criteria and techniques for learners' assessment and evaluation. The described benefits and limitations of each technique may facilitate the choice of the most adequate way to evaluate a particular SG. Despite the large scope of this survey, this work does not claim to include all existing techniques of learners' assessment and evaluation. However, it includes major themes identified in the literature, and provides a taxonomy where other works can be classified. An important research direction emerging from this research is the development of new implicit

assessment/evaluation approaches that consider both emotional and social dimensions in multi-actors SG for crisis management. These approaches can be embedded into SG to provide learners with relevant information about their emotional and social states and to improve training results.

## References

1. Walker, W.E., Giddings, J., Armstrong, S.: Training and learning for crisis management using a virtual simulation/gaming environment. *Cogn. Technol. Work* **13**(3), 163–173 (2011)
2. Oulhaci, A., Tranvouez, E., Fournier, S., Espinasse, B.: Improving players' assessment in crisis management serious games: the SIMFOR project. In: Bellamine Ben Saoud, N., Adam, C., Hanachi, C. (eds.) *ISCRAM-med 2015*. LNBIP, vol. 233, pp. 85–99. Springer, Cham (2015). doi:[10.1007/978-3-319-24399-3\\_8](https://doi.org/10.1007/978-3-319-24399-3_8)
3. Haferkamp, N., Kraemer, N.C., Linehan, C., Schembri, M.: Training disaster communication by means of serious games in virtual environments. *Entertainment Comput.* **2**(2), 81–88 (2011)
4. van, Ruijven, T., Igor, M., de Mark, B.: Multidisciplinary coordination of on-scene command teams in virtual emergency exercises. *Int. J. Crit. Infrastruct. Prot.* **9**, 13–23 (2015)
5. Carole, A., Franck, T., Etienne, D., Odile, P., Mira, T.: SPRITE - participatory simulation for raising awareness about coastal flood risk on the Oleron Island. In: International Conference on Information Systems for Crisis Response and Management in Mediterranean Countries, pp. 33–46 (2016)
6. Bellotti, F., Kapralos, B., Lee, K., Moreno-Ger, P., Berta, R.: Assessment in and of serious games: an overview. *Advances in Human-Computer Interaction* (2013)
7. Auferbauer, D., Berg, R.P., Hellingrath, B., Havlik, D., Middelhoff, M., Pielorz, J., Widera, A.: Crowdsourcing and crowdtasking in crisis management: Lessons learned from a field experiment simulating a flooding in the city of the Hague. In: International Conference on Information and Communication Technologies for Disaster Management (2016)
8. Shute, V.J.: Stealth assessment in computer-based games to support learning. *Comput. Games Instr.* **55**(2), 503–524 (2011)
9. Di Loreto, I., Mora, S., Divitini, M.: Collaborative serious games for crisis management: an overview. In: Proceedings of the IEEE International Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises, pp. 352–357 (2012)
10. Mora, S., Divitini, M.: Supporting debriefing with sensor data: a reflective approach to crisis training. In: Hanachi, C., Bénaben, F., Charoy, F. (eds.) *ISCRAM-med 2014*. LNBIP, vol. 196, pp. 71–84. Springer, Cham (2014). doi:[10.1007/978-3-319-11818-5\\_7](https://doi.org/10.1007/978-3-319-11818-5_7)
11. Pradeepa, T., Jean-Marc, L., Mathieu, M., Amel Y.: How to evaluate competencies in game-based learning systems automatically? In: Proceedings of International Conference on Intelligent Tutoring Systems, pp. 168–173 (2012)
12. Zagorecki, A., Johnson, D.E.A., Ristvej, J.: Data mining and machine learning in the context of disaster and crisis management. *Int. J. Emergency Manage.* **9**, 351–365 (2013)
13. Silva, V.S.R., Dargains, A.R., Felício, S.P.A.S., Souza, P.R.A., Sampaio, F., Motta, C.L.R., Borges, M.R.S., Gomes, J.O., Carvalho, P.V.R.: Stop disasters: serious games with elementary school students in Rio de Janeiro. In: International Technology, Education and Development Conference, pp. 1648–1659 (2014)
14. Angelo, T., Cross, K.P.: *Classroom Assessment Techniques a Handbook for College Teachers*. Jossey-Bass A Wiley Imprint, San Francisco (1993)