J Ambient Intell Human Comput (2018) 9:497–511
DOI 10.1007/s12652-016-0436-6

ORIGINAL RESEARCH

Decoupling assessment and serious games to support guided exploratory learning in smart education

Mohammad AL-Smadi[1] · Nicola Capuano[2] · Christian Guetl[3,4]

Received: 18 July 2016 / Accepted: 26 November 2016 / Published online: 19 December 2016

Abstract This research proposes an enhanced approach of decoupling assessment and serious games to support fire evacuation training in smart education. The proposed assessment approach employs an evidence-based dynamic assessment and feedback to guide players through school's building evacuation. Experimentation results show the applicability of the proposed assessment approach in enhancing fire evacuation training using serious games. Moreover, students were engaged to the proposed learning scenarios and their overall fire evacuation assessment were enhanced using the guided exploratory game-based training.

Keywords Dynamic assessment · Emergency training · Exploratory learning · Serious games · Fire evacuation

1 Introduction

In real-world fire evacuation, building environment and people behaviour are the main two factors that determine a fire evacuation success (Kobes et al. 2010). Building environment design has been covered in many research studies through simulating buildings evacuation. Examples on such simulations include both fixed evacuation paths such as EVACNET4 (Francis and Saunders 1979; Kisko et al. 1998; Zhang et al. 2009), EESCAPE (Kendik 1986), Exit89 (Fahy 1994), EGRESS (Ketchell et al. 1993, BGRAF, EXODUS, SIMULEX (Gwynne et al. 1999; Santos and Aguirre 2004), and dynamic evacuation paths and patterns such as FDS + Evac (Korhonen and Hostikka 2009), STEPS (MacDonald 2003), using agent-based elevator model (Tan et al. 2015), Pathfinder[1] and its application in fire evacuation (Ding and Yang 2013; Ding et al. 2015). However, designing real-world buildings perfectly to facilitate occupants evacuation is not enough to guarantee safe fire evacuation. Occupants need to be trained several times in order to shape their behaviour towards evacuating fire safely (Fahy and Proulx 2009; Kuligowski and Gwynne 2010; Ronchi and Nilsson 2013).

Fire evacuation training is getting more important in school education. However, how can educational systems shape pupils' behaviour during their schools' fire evacuation training? Moreover, how educational systems can provide school pupils with online learning activities that support fire evacuation training? Having an exploratory interaction type, serious games have become an important didactic resource for skills training and behavioural change (Williams-Bell et al. 2015; Girard et al. 2013; Chittaro and Ranon 2009; Ribeiro et al. 2012). Guided learning refers to the constructivist notion of learning where educators are facilitators assisting learners to make their own sense of learning content (Vygotsky 1986; Fosnot and Perry 1996). Learning with Sims enables aspects of discovery, where learners are required to infer knowledge from information given and this is where intuition can guide the construction of knowledge based on the prior knowledge and the

✉ Mohammad AL-Smadi
maalsmadi9@just.edu.jo

1 Computer Science Department, Jordan University of Science and Technology, Irbid, Jordan

2 Department of Information Engineering, Electric Engineering and Applied Mathematics, University of Salerno, Salerno, Italy

3 Graz University of Technology, Graz, Austria

4 Curtin University, Perth, Western Australia, Australia

1 http://www.thunderheadeng.com/pathfinder/

knowledge acquired from the given situation. However, in simulations and virtual worlds, learning objectives are frequently expressed less explicitly and less clearly defined (Swaak et al. 1998). Moreover, Research on learning processes with simulations have identified that learners may encounter problems when learning with simulations (Friedler et al. 1990; Glaser et al. 1992; Njoo and De Jong 1993; Veenman et al. 1997; De Jong and Van Joolingen 1998). Under established models relating cognition and learning (Bloom and Engelhart 1969), a danger of cognitive overload could be perceived from media rich environments. Through media rich environments, learner can extract high quantities of information in numerous ways (Swaak et al. 1998).Yet an exploratory (De Freitas and Neumann 2009), or experiential (Kolb 1984) approach to learning has been shown to work well in a simulator-driven learning environment (Raybourn 2007). Therefore, it could equally be argued that the 'ambient fidelity', e.g., the fidelity of the surrounding environment extraneous to the learning objectives, can also play a key role in sustaining the engagement of learners.

Serious games represent a challenging as well a rich domain for assessment practices. However, the efficacy of any assessment approach is highly related to the target demographic, usage context, choice of technology, and underlying pedagogy (De Freitas and Oliver 2005). Hence an attempt to evaluate any assessment model typically results with lack applicability when transferred to: other groups of learners, different context, and different educational situations. When players interact with a game they eventually take possible actions predefined in the game model of player–game interactions. This can be utilized to define assessment forms of monitoring players' activities based on logging all actions within the game session which can then be used to grade the player activities within the game. Therefore, serious games have to be carefully designed so that instructors can define assessment rules for specific state transitions during the game session (Burgos et al. 2008). Assessment and feedback should be seamlessly integrated into games without compromising gameplay. This means that assessment and feedback should be provided non-invasively without interrupting player's (learner) behaviour during the game (Kickmeier-Rust et al. 2009).

When games are designed, game mechanics, game dynamics and game aesthetics (MDA) can be considered as a framework to have a well-designed game (Hunicke et al. 2004). Game mechanics are related to possible game interactions and controls (i.e., points, levels, challenges, etc.), game dynamics are related to the resulted action or consequences of mechanics (i.e., rewards, achievements, status, self-expression, etc.), whereas game aesthetics is the felt user experience when playing the game (i.e., curiosity, satisfaction, fun, etc.) (Bunchball 2010; Deterding et al. 2011; AL-Smadi 2015). However, the process of evaluating a specific player-game interaction (game mechanics) and providing a consequent game action (game dynamics) can be seen as embedded assessment in serious games. This process of evaluation is designed as game 'checkpoints' where rules are mainly used to evaluate player interaction and provide game dynamics. Moreover, evaluation process is embedded within the game and cannot be changed during game runtime. Therefore, the application of games in education is static and limited to the context and educational settings the game has been designed for.

To this end, this research extends the work presented in (AL-Smadi et al. 2012) through proposing an enhanced flexible design for a dynamic assessment approach for serious games and 3D immersive environments. In addition to technical enhancements, the assessment model is extended to use a semantic knowledge base that represents abstract spatial location of the environment (the game)objects and avatars, observable state of entities (items and avatars), and possible interactions between objects and avatars. Moreover, an extensive literature review has been conducted and is presented as an extension of previous work. The proposed approach was evaluated in real learning settings to support guided exploratory learning in fire evacuation training. Results show that the approach is applicable for emergency training using serious games.

Proposed approach supports both types of summative and formative assessment and is fully integrated within the Learning Management System (LMS) for both assessment authoring and delivery. Moreover, in contrast to related works, the proposed approach advocates the decoupling of the assessment design and the game engine in order to have a flexible and interoperable assessment in serious games, where instructors can easily activate and deactivate assessment rules (games 'checkpoints') to evaluate and guide learners through the game scenarios without changing game design or components. Moreover, this decoupling fosters serious games to be easily extended to support different contexts and variety of smart educational settings.

The rest of this paper is organized as follows: Sect. 2 sheds the light on related works for assessment in serious games, Sect. 3 discusses the proposed approach by focusing on the integrated dynamic assessment architecture and scenarios, Sect. 4 presents the approach evaluation and results, Sect. 5 discusses the results and findings, and finally, Sect. 6 concludes this research.

2 Assessment in serious games

Assessment in serious games can be classified into: (1) Micro-adaptive and non-invasive assessment (integrated assessment), (2) Players' state-based assessment (embedded assessment), and (3) quest-based assessment (invasive assessment). Hainey et al. (2014) distinguish between embedded and external assessment in serious games as, embedded assessment is part of learning activity (is part of the serious game) (Underwood et al. 2010), has a formative purpose (Shute et al. 2009), sometimes used for 'micro-adpativity' (Kickmeier-Rust and Albert 2010), and should not interrupt the game flow (Eseryel et al. 2011), whereas external assessment is not part of the serious game and in many cases interrupt the game flow (Eseryel et al. 2011). Integrated assessment is the same as embedded assessment. However, integrated assessment is delivered within the game and the assessment logic (engine) is external to the game (part of the assessment module or the LMS), whereas in embedded assessment both assessment logic and delivery are embedded within the game. However, integrated and embedded assessment are used interchangeably in literature.

2.1 Micro-adaptive and non-invasive assessment (integrated assessment)

Shute et al. (2009) discuss the evaluation of players' progress within immersive games via what they called stealth assessment—embedded formative assessment within the immersive game—based on extending the evidence-centered design assessment model (ECD) (Almond et al. 2002) with an action model instead of task model which can be used by Bayesian networks to track player actions within the game and provide an evidence of progress. However, what they discussed represents a summative approach by which they evaluate the progress of the player in terms of interactions against the evidence model—part of ECD—to provide an evidence of learning and skills achievement after playing the game.

Kickmeier-Rust et al. (2009) propose the so-called micro-adaptivity approach for assessment in educational games. The approach has been developed in the context of the Enhanced Learning Experience and Knowledge TRAnsfer (ELEKTRA[2]) project. The ELEKTRA framework uses the Competence-based Knowledge Space Theory (CbKST) to model the competencies required by the student to achieve a learning goal. The basic idea of CbKST is to associate problems in a domain with skills in order to provide a model of competencies for learning. The ELEKTRA game and its successor 80Days[3] game tracks the player interactions and uses them to update the players' competence state represented in the CbKST for the learning. However, according to Kickmeier-Rust and Albert (2010) the approach demands extra load on authoring aspects to define all required information for the models as well as computational load as the game updates the CbKST based on each player action.

2.2 Players' state-based assessment (embedded assessment)

Moreno-Ger et al. (2008) propose the adventure game engine ($< e-Adventure >$). In $< e-Adventure >$, assessment rules are defined using Extensible Markup Language (XML) which acts as a game model. This model contains information about the game storyline, scenes, characters, their associated resources, as well as assessment and adaptive rules. When a player is interacting with the game an associated assessment rule is triggered based on internal flags associated with specific states of the game. Moreover, the authors claims that the $< e-Adventure >$ is flexible to be used with educational modeling languages, i.e., IMS Learning Design (IMS-LD)—to design the pedagogical impact of using assessment rules in the game engine. However, $< e-Adventure >$ is a 2D simulation environment with assessment model that lacks services to be used in other gaming environments.

JDoc (Junior Doctor) (Sliney and Murphy 2008) presents junior doctors with virtual patients and evaluates their progress towards the final diagnosis and prescription. The doctor's final diagnosis and prescription are then reported to a senior doctor for summative assessment. JDoc is also able to evaluate doctors' states of diagnosis and provide formative feedback based on their decisions (Sliney et al. 2009; Sliney and Murphy 2011).

Focusing on behavioural assessment in 3D virtual worlds, there are proposals to use assessment approaches from serious games in which player interactions are tracked and used to evaluate their behaviour and skills based on artificial intelligence methods such as decision trees, finite state machines, and Bayesian networks (Ibáñez et al. 2010). However, these approaches also lack to some extent flexible teacher control over the assessment method and are not applicable for different contexts and learning scenarios.

Maderer et al. (2013) presents a dynamic assessment for avatars behaviour in Open Wonderland. The approach extends the work of AL-Smadi et al. (2012) by using a semantic knowledge base that describes abstract spatial location of items and avatars, abilities and observable state of entities (items and avatars), and possible interactions between objects and avatars. This knowledge base is used by the assessment module to invoke events based on avatars (users) interactions and to provide feedback.

[2] http://www.elektraproject.org

[3] http://www.eightydays.eu

2.3 Quest-based assessment (invasive assessment)

Connolly et al. (2011) present a game called 'Tower of Babel ARG. Tower of Babel ARG uses quest-based assessment integrated to a digital story. Players have to socially interact and collaborate into learning a foreign language. The game provides a type of formative assessment based on Alternate Reality Game (ARG) to solve problems in foreign language learning. In the same direction, the 'Requirements Collection and Analysis Game (RCAG) game (Hainey et al. 2011) where players have to collaboratively collect requirements and analyse them. Peer-assessment is provided to evaluate players' progress on collecting and analysing requirements.

A situated learning (Lave and Wenger 1991) based serious game was developed by the Scottish Qualification Authority (SQA) to foster vocational training and qualification (McAlpine et al. 2010). The game places a trainee into a workplace similar environment and provides formative and summative assessment (questioning based). The game produces a report based on players' results of the game-based assessment.

A Web-based multiplayer online role playing game (MORPG) for assessing students' Java programming skills is presented in (Chang et al. 2010). In addition to coding quests, the game provides other quests such as look for NPCs, single/multiple-choice quests, and fill-in-the-blank quests.

Regarding providing assessment in immersive 3D environments. The Sloodle project (Simulation Linked Object Oriented Dynamic Learning Environment) enables Moodle integrated quizzes within Second Life using Quiz chair tool (Kemp et al. 2009). QuizHUD (Heads Up Display) (Bloomfield and Livingstone 2009) enables avatars in Second Life to touch items and complex objects in order to provide response while answering regular multiple choice questions. However, both Sloodle Quiz chair and QuizHUD are limited to simple types of questions such as multiple choice questions and lack the support of other 3D environments—such as Open Wonderland and OpenSim or serious games.

Literature reviews on feedback in digital educational games (Shute 2008; Mory 2006) highlight the importance of formative models for feedback provision. Dunwell et al. (2011) discuss feedback aspects in digital educational games based on Roger's classification (Rogers 1951) into: *evaluative* (players get a score), *interpretive* (players get a score and the wrong action), *supportive* (players get a score and guidance information), *probing* (players get a score and analysis of why the player did the wrong action), and *understanding* (players get a score and analysis of why the player did the wrong action as well as guidance for supportive steps or learning material) forms. Moreover they propose a four-dimensional approach for feedback provision in serious games. According to their approach the following aspects should be considered: (1) *Type*: feedback type differs based on Roger's classification of feedback—discussed above—with respect to students, teachers, or technology thus required aspects to classify feedback such as measure variables, their relationships model, learner model, knowledge model, and domain model- should be considered. (2) *Content*: content can be classified with respect to the learning outcomes into essential or desirable. (3) *Format*: the media used to represent feedback (e.g. text, image, voice, etc.), and (4) *Frequency*: the rate of feedback provision to students differs with respect to instructors, technology, pedagogy, and learner preferences control. Hence, feedback can be *immediate*, *delayed*, or *dynamic* based on the domain and learner action type.

Players make decisions when they interact with the game by taking right/wrong actions and paths. The game platform should have the possibility to define checkpoints (assessment rules) so that to assess players' interactions and decisions. Moreover, it should provide valuable feedback. According to Burgos et al. (2008) in order to have an adaptive and personalized serious game within the domain of game-based learning, the game engine has to be integrated with a LMS. Accordingly, our research assumes that (1) the LMS uses the Log of player interactions within the game session to provide more personalized and adaptable content, (2) players' flow within the game will form like a learning path where a third-party tool is needed to, interact with the game engine, retrieve player's progress, and communicate this progress with the LMS in order to evaluate player interactions, and update her learning profile. Moreover, (3) adaptive and personal game content can be provided during the game next phases based on player's progress.

To this end, assessment approaches in serious games and 3D environments can be classified into (1) embedded assessment: through which assessment rules and consequent feedback is scripted in the game scenarios. This form of assessment is static, and does not consider learning styles and learner preferences, (2) dynamic flexible assessment: in which assessment acts as a middleware between the game engine and the LMS. The dynamic and flexible assessment is authored separately of the game design and is done by the teacher. The dynamic flexible assessment supports both summative and formative assessment where user's interactions with the game can be evaluated in a summative way of assessment and used to update the learner profile within the LMS. Whereas in the formative one, dynamic assessment and guidance are used to scaffold learner behaviour while playing and exploring the game. Next section discusses how our proposed approach employs the second type (i.e., dynamic flexible assessment).

3 Evidence-based dynamic assessment and feedback in serious games

3.1 The game

The game—which was developed at the Serious Games Institute (SGI) at Coventry University—adopts a freely navigable 3D environment, created within the Unity Engine.[4] The game contains elements of crowd simulation within fire evacuation scenarios, effectively placing the player within the building and monitoring their actions as they evacuate. Hence, provide assessment and effective feedback, it is essential that the game monitors and correctly identifies key actions which may indicate correct and incorrect behaviours. The principal means through which it is proposed is achieved through the implementation of virtual checkpoints within each scenario, recording players' time and state as they pass within a radius of a single point within the virtual space. Nevertheless, the game designer annotates the pedagogical objects (game object that teachers may use to provide feedback and guidance) and share these annotations as XML file with the LMS. The annotations are used by the assessment designer to provide assessment rules associated with feedback to the player once an event regarding one of these pedagogical objects is invoked and sent to the Assessment Engine (see Sect. 3.3 for technical details).

3.2 Assessment model

Findings from literature stress on the importance of having flexible, dynamic and integrated assessment forms for serious games and 3D immersive environments. In order to achieve these requirements, the proposed assessment approach adopts the Evidence-centered assessment design (ECD) framework (Almond et al. 2002). ECD is a framework that explains the structures of assessment arguments, their elements and process, as well as the interrelationships among them. ECD provides a conceptual assessment framework (CAF) which acts as an intermediate step between the output of the domain analysis and domain modelling steps. CAF specifies the knowledge and skills to be assessed, conditions for assessment and evaluations, as well as type of evidences for the provided tasks assessment. Moreover, ECD has an assessment implementation step that describes the requirements for processes during the assessment delivery phase (Mislevy and Riconscente 2005). CAF discusses the substantial, statistical and operational aspects of assessment elements. Moreover, it covers technical details such as, specifications, operational requirements, statistical models, details of rubrics. CAF consists of a set of modules which provides specifications to answer critical questions such as:

1. What Are We Measuring: The Student Model
2. How Do We Measure It: The Evidence Model
3. Where Do We Measure It: The Task Model
4. How Much Do We Need to Measure: The Assembly Model
5. How Does It Look: The Presentation Model.

Moreover, these five models describe the requirements for the objects in the assessment delivery system. The Delivery System Model describes the collection of student, evidence, task, assembly, and presentation models necessary for the assessment and how they will work together (Mislevy et al. 2004).

ECD based assessment adopts the four-process architecture of providing assessment (Almond et al. 2002): activity selection, presentation, response processing, and summary scoring. The creation of the assessment task starts by the activity selection process where the administrator (instructor) selects and sequences tasks from the task/evidence composite library (a database of possible tasks, their description, materials, rules, and evidence parameters). Then information is sent to the presentation process, which delivers the assessment task to the participant (student). Relevant materials can be retrieved from the task/evidence composite library for instance, assessment paper (traditional assessment) or images, audio/video files (e-assessment). The presentation process records the students responds as a work product which can be assessment paper script, or computer file and then delivers this work product to the response processing section for evaluation. The evaluation process may consist of simple scoring process or more complex series of evaluation for the students responses. The evaluations are then passed to the summary scoring process which updates the scoring record. The scoring record contains all the judgments about students' knowledge, skills level, and abilities based on pre-defined evidences provided for all tasks. According to Almond et al. (2002), separating the response processing step from both summary scoring and presentation is vital to an evidence-based assessment design. Moreover, such separation supports the reuse of the assessment task in multiple contexts. Two types of feedback can be delivered based on this architecture: (1) *task-level feedback*, which represents the immediate feedback based on student responses independently of evidence from other tasks, and (2) *summary feedback*, which reports the accumulated observations from

[4] http://unity3d.com

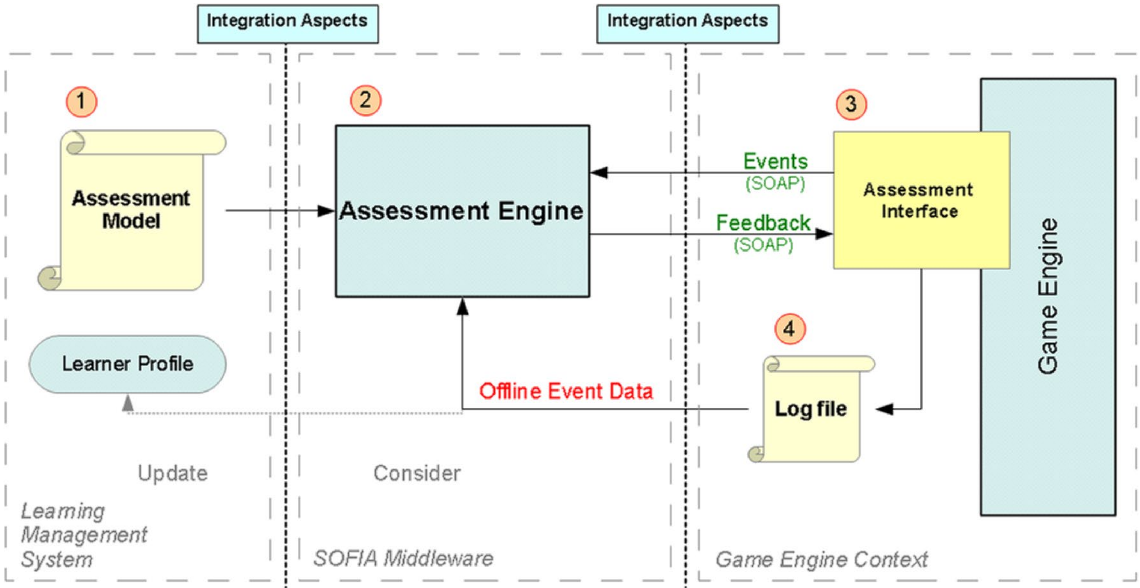

Fig. 1 Enhanced approach for decoupling assessment and serious games

the scoring record based on tasks' evidences to the participant (student).

3.3 Dynamic assessment and feedback

Providing assessment for serious games holds some challenges especially when it comes to provide formative dynamic assessment and feedback (AL-Smadi et al. 2012; Shute et al. 2009; Kickmeier-Rust et al. 2009). In order to tackle this problem an enhanced approach for integrated dynamic assessment in 'stealth' mode (i.e., none invasive) has been developed through the architecture depicted in Fig. 1. The architecture is designed to implement the ECD four-process architecture of providing assessment where instructors are enabled to select a set of assessment rules out of the task/evidence models (Where/How do we measure a specific objective) and the assessment architecture will use these assessment rules within the game to evaluate and guide players' progress. The game (presentation layer) records player's responses to the assessment rules and send them back to the LMS. The LMS handles the response processing, provides feedback (formative assessment), and updates the scoring record (summative assessment). Provided feedback follows the ECD feedback types: (1) *task-level*, which is provided immediately to players based on their right/wrong interactions, (2) *summary feedback*, which reports the summative evaluations of players' interactions.

As a result, the developed scenario to foster integrated dynamic assessment for game-based learning (GBL) has the following components (see Fig. 1):

1 *Assessment model* is an XML based description of behavioural patterns and associated consequences (see Fig. 2). Behavioural patterns are defined through sequences of possible player actions and conditional matches. Whereas, consequences have the primary goal of providing feedback— action-based guidance— to the player within the game engine after detecting specific pattern by the assessment engine. Consequences can take a form of game-design actions to enable game internal measurement operation (e.g., stop watches, gain/loss of points, etc.) and/or feedback messages that are presented to the player in order to guide her progress within the game. The assessment model is authored by the teacher using assessment module within a LMS that is designed to interface with assessment model. The model is then retrieved by the assessment engine with respect to the learning task. Following the work of Maderer et al. (2013), the assessment model is extended to use a semantic knowledge base that represents abstract spatial location of items and avatars, observable state of entities (items and avatars), and possible interactions between objects and avatars. This knowledge base is used by the assessment engine to invoke events based on avatars (users) interactions and to provide feedback.

2 *Assessment engine* is developed as Web services within the middleware of the Service-Oriented Flexible and Interoperable Assessment framework (SOFIA) (AL-Smadi and Guetl 2011). As part of the dynamic assessment and feedback scenario, the assessment engine uses the related assessment model. The retrieval of

Fig. 2 Part of the assessment model for the event of a player collection of her bag during fire evacuation

```
<behaviour-pattern>
  <match-action name="UseObject">
    <compare type="Equal">
      <get property-name="object_type" />
      <value type="String">bag</value>
    </compare>
  </match-action>
  <consequences>
    <call name="Console.DebugMessage">
      <parameter name="message">
        <value type="String">
          You took your bag before leaving the class room.
          You should not collect your possessions before evacuating.
        </value>
      </parameter>
    </call>
  </consequences>
</behaviour-pattern>
```

the assessment model is based on the learning task as discussed earlier. Using the assessment model, the assessment engine analyses and matches all possible assessment rules when invoked run-timely by the game engine through receiving the game flow events, and provides task-level feedback (as defined in the task assessment model). Whereas, in the post evaluation scenario, the assessment engine uses the game log file to score and report the overall player's progress to the LMS. Moreover, players' results are used to update the learner profile in the LMS.

3. *Assessment interface* handles the communication between the game engine and assessment engine. The assessment engine is managed through Web services developed as part of SOFIA middleware (AL-Smadi and Guetl 2011). For these Web services interfaces are provided and used to interact with the assessment engine methods. The services are described using the Web Services Description language (WDSL) and use the Simple Object Access Protocol (SOAP) for messages communication and transport.
4. *Log file* is created by the game engine to track players' interactions and environment changes which are logged in an XML-based log file (see Fig. 3). The log file is used for post evaluation scenario to provide a report based on player behaviour and performance within the game environment.

3.4 Assessment scenarios

The proposed assessment architecture supports two main scenarios for game-based learning assessment:

1. *Summative assessment (post evaluation)* in which a log-file has been designed to hold all the actions related to the assessment scenario for specific context e.g. fire evacuation training through tracking the players' interactions. Moreover, an assessment engine is developed to interact with an assessment model to evaluate the players' progress—represented by log file actions—against a predefined assessment rules to assess specific learning objectives, e.g., crawling in smoke during fire evacuation.
2. *Formative assessment (dynamic assessment and feedback)* in which an assessment interface is attached to the game engine in order to handle run-time events out of players' interactions and call the assessment engine to evaluate those actions based on predefined assessment rules in the assessment model. Then predefined feedback associated to those assessment rules is dynamically provided to the player (task-level feedback). The rate of feedback provision to players depends on instructors, technology, pedagogy, and learner preferences. Hence, feedback can be immediate, delayed, or dynamic based on the domain and learner action type. This has been considered in the design of the feedback block in the assessment model in order to support both scenarios for assessment.

```
<logfile>
    <event time="00:00:19.6811257" player-id="noname">
        <action name="UseObject">
            <set property-name="object_type">bag</set>
        </action>
    </event>
</logfile>
```

Fig. 3 Part of the log file related to the event of a player collected her bag during fire evacuation

The application of the proposed dynamic assessment and feedback scenario consists of two main phases:(1) *assessment and feedback definition phase*: in which the educator defines assessment rules based on possible interactions of the player with the game pedagogical objects (e.g., bag, elevator) tagged during the game design phase. Figure 2 shows an example of part of the assessment model that was used to train students not to collect their possessions (i.e., bag) during fire evacuation, and (2) *gameplay phase*: where the assessment rules are invoked based on players' interactions with the game. The Game engine tracks players' behaviour and environment changes and save them to a log file designed for the game-based learning(GBL) assessment. Players' interactions with game tagged pedagogical objects (e.g., bag, elevator) invokes an event to save a record to the log file. A JavaScript function is developed to send events from the game engine to the assessment engine. Invoked events are then evaluated by the game engine and feedback is provided to players.

For instance, a scenario for teaching students that they should not collect their possessions during fire evacuation was tested. The instructor used the GBL-Assessment module (as part of the LMS, see Fig. 1) and designed an assessment rule for the game object 'bag which was added automatically to the assessment model. As presented in Fig. 2, the created assessment rule consists of: *match action* ('Use-Object), *value* ('bag), and *consequences*(name: 'Console. DebugMessage, parameter: 'message, value: 'You took your bag during fire evacuation! You should not collect your possessions before evacuating).

During gameplay, the player collected his bag before leaving the class room. This action fired an event and saved it to the log file (see Fig. 3). The JavaScript function in the Web-based game player called the assessment engine web services with the logged event. The assessment engine evaluated that event against the assessment rule defined by the instructor (Fig. 2) and replied with the feedback message. The feedback message was then presented to the player using a NPC at the bottom of the screen (see Fig. 4, message in Green). The message on top of the screen (see Fig. 4, message in Red) was generated by the embedded assessment within the game (i.e., using the game embedded script). This assessment scenario was used to guide players not to collect their possessions during fire evacuation using dynamic feedback provision.

The decoupling of the game engine—as a complex learning resource—and the assessment engine—utilized via web services—fosters the accommodation of various learning contexts and pedagogical approaches. For instance, the same integrated assessment approach was used to provide guidance in other scenarios such as, using elevators during fire evacuation as a prohibited action with a feedback of the type 'game-design action (i.e., loss of

Fig. 4 The player action represented by collecting his possessions triggered an assessment rule which has a consequence of providing feedback using a non-player virtual character

players' points—Game Over), and crawling in a smoky area as correct behaviour with a 'supportive feedback message of 'Well Done, keep crawling until you reach a safe area. Using elevators during fire evacuation is a debatable case, some recent studies showed that using elevators during fire evacuation especially for disabled people is a need (Bukowski 2012). In high-rise buildings (such as the Burj Khalifa 800m high) using the stairs for fire evacuation is fatigue (Andrée et al. 2015). Survivals of 9/11 reported that they had to stop and rest several times during evacuation (Averill et al. 2005). Using the flexibility of our integrated assessment approach, the instructor is able to activate/deactivate the assessment rule on using elevators during fire evacuation based on the pedagogical approach and context.

Focusing on the post evaluation assessment scenario, the summative assessment (post evaluation) output can be used to update the player's competency level, adapt the game to the player level of competence, as well as provide summary feedback regarding the player's progress within the game.

4 Experimentation and evaluation

An experiment was undertaken with the participation of 4 high schools coming from the Campania region of Italy to validate the developed serious game and the integrated dynamic assessment approach. The experiment involved 45 voluntary students and 4 tutors: 8 students coming from schools 1, 3 and 4, 21 students coming from school 2, one tutor per school. The experimental sessions were organized in the PC lab of each school and spanned two hours each, two sessions were organized for school 2 with two subgroups of 10 and 11 students.

The experiment aimed at answering the following research questions:

- RQ1: To what extent the proposed approach is engaging and attractive for the student (*perceived value*)?
- RQ2: To what extent the adopted interface is friendly for the student (*usability*)?
- RQ3: To what extent playing with the game results in a real improvement of the student's fire evacuation skill (*effectiveness*)?

The following subsections summarize the applied experimental methodology as well as the obtained results and findings. A subset of the experimental data used in this paper has been also used in (Capuano and King 2015) with the aim of demonstrating the overall effectiveness of the serious game approach for emergency preparedness training rather than focusing on the underlying methodological approach (summarized by the three research questions described so far) as in this case.

4.1 Apparatus and stimuli

A questionnaire based approach was adopted to measure both the perceived value and the usability (with respect to the first two research questions), while the effectiveness (third research question) was evaluated by analysing users' performances and by interviewing the involved tutors.

In particular, to measure the level of engagement and attractiveness conveyed by the proposed learning approach in terms of perceived value, the participating students have been asked to provide an answer to the following set of questions:

1. How responsive was the game to actions that you performed?
2. How much did the graphical aspects of the game involve you?
3. How compelling was the experience of moving through your own school building?
4. How much did the hardware quality interfere or distract you from performing the assigned tasks?
5. To what extent have you been engaged in the solution of the game?
6. How easily did you control the interaction with the game?
7. Were the provided supporting learning resources consistent with correct and wrong actions performed within the game?

The questions are thought not only to evaluate the overall perceived level of engagement (question 5) but also to understand to what extent specific elements positively or negatively affect such engagement. Among the positive elements, we considered the graphical 'immersive' approach (question 2) or the fact that the students move in their own school (question 3). Among the possibly negative elements, we measured the lack of responsiveness of the game (question 1) and the lack in controlling the interaction (question 6), both impacting fidelity and, possibly, caused by low hardware quality (question 4). Finally, question 7 is specifically provided to evaluate the effectiveness (from the student perspective) of the integrated dynamic assessment approach.

The possible answers for these questions, corresponding to the five points of the Likert scale, are: *not at all* (1), *a little* (2), *moderately* (3), *very* (4) and *completely* (5). It is worth noting that, while the majority of questions have a positive polarity, question 4 has a negative polarity (i.e., a high score corresponds to negative evaluation).

To measure the usability of the developed prototype system we adopted the System Usability Scale (SUS) defined in (Brooke et al. 1996). SUS is based on a set of 10 questions whose answers are given on the five-point Likert scale: *strongly disagree* (1), *disagree* (2), *neither/nor* (3), *agree* (4) and *strongly agree* (5). The 10 items that compose the SUS questions are:

1. I think that I would like to use this system frequently;
2. I found the system unnecessarily complex;
3. I thought the system was easy to use;
4. I think that I would need the support of a technical person to be able to use this system;
5. I found the various functions in this system were well integrated;
6. I thought there was too much inconsistency in this system;
7. I would imagine that most people would learn to use this system very quickly;
8. I found the system very cumbersome to use;
9. I felt very confident using the system;
10. I needed to learn a lot of things before I could get going with this system.

Table 1 Detailed results obtained on questions about perceived value

Question	School 1		School 2		School 3		School 4		Overall	
	M	SD	M	SD	M	SD	M	SD	M	SD
Q1	2.00	0.76	2.67	0.66	3.63	0.74	3.38	0.74	2.84	0.96
Q2	3.00	0.93	3.33	0.73	3.38	0.52	3.25	0.71	3.27	0.86
Q3	3.25	0.71	4.10	0.77	3.75	0.71	3.88	0.64	3.84	0.95
Q4	4.25	0.71	3.81	0.81	2.38	0.52	2.63	0.52	3.42	1.10
Q5	3.13	0.64	4.00	0.84	4.63	0.52	4.75	0.46	4.09	1.05
Q6	3.50	0.93	3.14	0.65	4.00	0.76	3.63	0.74	3.44	0.93
Q7	3.75	0.71	4.00	0.77	3.75	0.71	4.13	0.83	3.93	0.94

SUS produces a single number representing a measure of the overall usability. To calculate the score, it is necessary to sum the score contributions from each item. For items 1, 3, 5, 7 and 9 (that have positive polarity) the contribution is the scale position minus 1. For items 2, 4, 6, 8 and 10 (that have negative polarity), the contribution is 5 minus the scale position. The final score, ranging from 0 to 100 is obtained by multiplying the sum of scores by 2.5.

To measure the *effectiveness* of the serious game with integrated assessment as a didactic resource, we collected, for each student and classroom, the average number of accesses to the game and the results of the assessment for all involved skills. To understand if the practice with the game helps in obtaining better performances on such skills, the *Pearson* correlation index between the number of accesses to the game and the overall student evaluation has been calculated. Interviews with the 4 involved tutors have also been conducted with the aim of obtaining additional advice.

4.2 Evaluation scenario

The experimental sessions were organized as follows. During the first 30 min, the purpose of the experiment was explained by tutors and questionnaires to be filled were distributed in a sealed envelope (students were not allowed to read them before the end of the experiment). Then the students were left free to familiarize with the learning environment without accessing the game. During the subsequent 60 min the students were allowed to access the game as well as connected learning resources arranged by the system automatically, according to students' performance.

The tutor stayed in the PC lab to oversee the work of the students while a technician, involved in the implementation of the game, was in the same room to provide assistance in case of need. The students were left free to exchange information and opinions between themselves and with the tutors. During the last 30 min, students filled out the questionnaires that were then collected by the tutor himself. Quantitative data needed for the measurement of the effectiveness were collected once for all the schools, directly on the game server (out of the LMS log files).

4.3 Results and findings

As presented in Table 1 the results obtained in the *perceived value* section of the questionnaire are promising. Students provided high average scores to questions 5 (M 4.09, SD 1.05) and 7 (M 3.93, SD 0.94), moderately high average scores to questions 2 (M 3.27, SD 0.86), 3 (M 3.84, SD 0.95), 4 (M 3.42, SD 1.10), and 6 (M 3.44, SD 0.93), and moderately low average score to question 1 (M 2.84, SD 0.96).

Table 2 summarizes the results obtained in the *usability* section of the questionnaire. The mean SUS score obtained for the serious game, over all users, according to provided answers is 61.28 (SD 4.35) with a minimum of 60.63 (SD 5.30) obtained in school 1 and a maximum of 63.44 (SD 3.76) obtained in school 4. The lower score obtained by schools 1 and 2 can be related again with the low performances of graphic hardware. The minimum individual SUS score (i.e., calculated basing on the input of a single student) is 52.50 (obtained in school 1) while the maximum is 70.00 (obtained in school 3). The box plots of Fig. 5 provide additional information about the distribution of overall scores with respect to both questionnaire sections.

With respect to the evaluation of the *effectiveness* of the serious game integrated with flexible assessment forms as a didactic resource, Table 3 reports the average number of accesses to the game and the results of the assessment mediated over all involved skills and normalized between 0 and 10. As it can be seen, the average evaluation for emergency skills is 6.44 (SD 1.65) that is 1.44 points above the passing threshold that was set to 5. The total number of accesses made by the 45 students is 173 with an average of 3.84 accesses per user (SD 1.81), a minimum of 1 and a maximum of 10. To understand if the practice with the game helps in obtaining better performances, the *Pearson* correlation index P_{XY} has been calculated between the number of accesses to the game by a student (X) and the overall evaluation obtained by the same student on

Table 2 Detailed results obtained on questions about usability

Question	School 1		School 2		School 3		School 4		Overall	
	M	SD	M	SD	M	SD	M	SD	M	SD
Q1	2.38	0.52	2.95	0.67	2.88	0.35	2.75	0.46	2.80	0.59
Q2	2.63	0.92	2.57	0.51	2.63	0.74	2.63	0.74	2.60	0.65
Q3	3.63	0.92	3.38	0.50	3.38	0.92	3.50	0.53	3.44	0.66
Q4	2.75	1.04	2.71	0.85	2.88	0.83	2.63	0.52	2.73	0.81
Q5	4.00	0.93	4.05	0.80	3.88	0.64	4.00	0.53	4.00	0.74
Q6	3.00	0.76	2.14	0.57	2.25	0.71	2.25	0.46	2.33	0.67
Q7	2.63	0.52	2.57	0.51	2.75	0.46	2.75	0.46	2.64	0.48
Q8	2.00	0.76	2.52	0.51	2.00	0.53	2.25	0.71	2.29	0.63
Q9	4.13	0.83	4.24	0.62	4.50	0.76	4.75	0.46	4.36	0.68
Q10	2.13	0.64	3.10	0.62	2.75	0.71	2.63	1.19	2.78	0.82
SUS	60.63	5.30	60.36	4.20	62.19	4.11	63.44	3.76	61.28	4.35

Table 3 Correlation between the number of accesses and the assessment results

Subset	Accesses			Assessment		Accesses/ Assessment Correlation
	Total	M	SD	M	SD	
School 1	33	4.13	1.25	6.63	1.19	0.71
School 2	86	4.10	2.28	6.19	1.86	0.76
School 3	27	3.38	1.19	6.75	1.28	0.63
School 4	27	3.38	1.41	6.63	1.92	0.75
Overall average	173	3.84	1.81	6.44	1.65	0.69

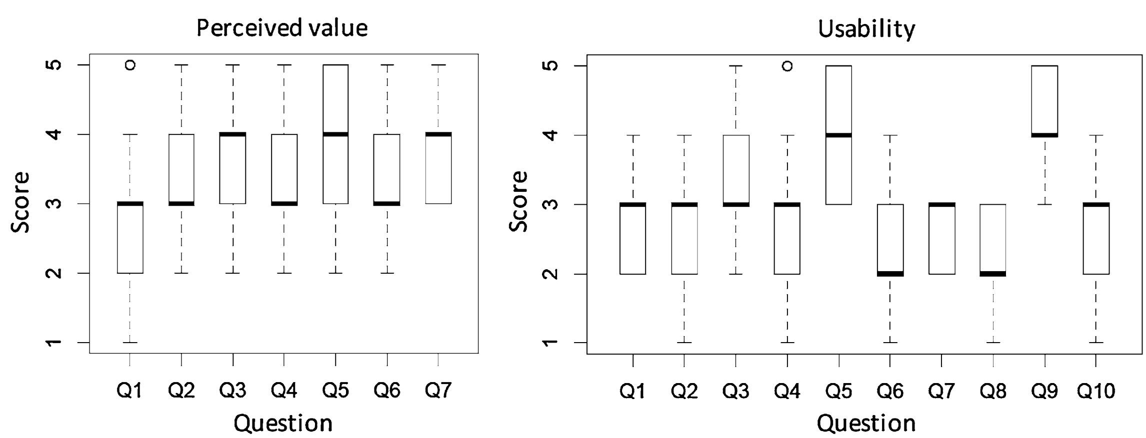

Fig. 5 Mean scores obtained on questions about perceived value (*left*) and usability (*right*)

emergency skills (Y). The obtained result of (P_{XY} 0,69) shows an almost strong direct correlation between the two variables (where a correlation is usually defined as strong when P_{XY} is greater than 0,7).

Some interviews with the 4 tutors were also performed with the aim of obtaining additional advice about the effectiveness of the proposed approach. The tutors unanimously consider the serious game as a valuable learning resource for emergency preparedness. They found the majority of students very participatory during the experiment and noticed an improvement on their skills. Three of the interviewed tutors suggested the inclusion of the game in the standard school curricula while one of them also proposed to extend the experiment to teachers and staff members. As for negative aspects, two tutors over 4 suggested to improve the usability of the game that, on some older PCs, presented a low fluency making difficult and unsmooth interaction.

5 Discussion

In order to evaluate the serious game with evidence-based integrated dynamic assessment as a didactic resource, the results coming from the questionnaires on perceived value and usability as well as the results of the effectiveness analysis are used.

Regarding the perceived value (RQ1), in particular, students assigned a high average score of 4.09 to question 5 (SD 1.05) showing that the proposed training method was compelling and engaged the learners. The average score of 3.27 assigned to question 2 (SD 0.86) shows that the visual aspect is only partially responsible for the high level of engagement. Other explanations could be found in the game playability and in the virtualization of the own school buildings of participating students. This latter hypothesis is supported by the average score of 3.84 assigned to question 3 (SD 0.95). Moreover, the average score of 3.93 assigned to question 7 (SD 0.94) shows that the provided feedback, based on the integrated dynamic assessment approach for students' correct/wrong interactions with the game, was perceived as valuable and consistent. This sustains the effectiveness of the proposed evidence-based approach.

The moderately low average score of 2.84 assigned to question 1 (SD 0.96) could be justified by taking into account that the PCs that have been used in the experiment, belonging to school labs, did not have any gaming dedicated graphics hardware. Therefore, in many cases the game was played with a low frame-rate. This hypothesis is also confirmed by the relatively high average score of 3.42 obtained on question 4 (SD 1.10). It is worth noting that the lowest values for question 1 were obtained by the same schools (1 and 2) that obtained the highest score on question 4. Eventually, the average score of 3.44 obtained on question 6 (SD 0.93) suggests that the user interaction with the game is good but can be improved. Even if this result may be affected by the hardware performance issues, it can be acceptable by considering the prototypical nature of the system.

Regarding the proposed approach usability (RQ2), beyond the small differences among schools, the overall SUS result of 61.28 (SD 4.43) is encouraging given the prototypical nature of the system. The lower score obtained by schools 1 and 2 can be related again with the low performances of their graphic hardware. However, focusing on the usability of the integrated assessment and despite the decoupling of the assessment and the game engines, students rated the integrated functionality (question 5) provided by the game (including the run-time assessment and feedback) with a high average score of 4.00 (SD 0.74), and the inconsistency of the game components and functionality (question 6) with a low average score of 2.33 (SD 0.67). This indicates that the integrated dynamic assessment provided between the game engine and the learning environment is usable and stable. Moreover, this also proves that the decoupling of the assessment engine (integrated with the learning environment) and the game engine is applicable and achieves positive user experience.

Regarding the effectiveness of the serious game integrated with flexible assessment forms as a didactic resource (RQ3), the almost average results of the evaluation for emergency skills of 6.44 (SD 1.65) can be explained by the nature of the learning activity as students were enabled to access and repeat the game several times in a way to evacuate a fire. Each time students play the game, they learn the fire evacuation rules based on guided exploratory approach. Moreover, their overall assessment after each try is used to update their learning profile. Therefore, the average of student assessment for the fire evacuation skills is based on all the tries not on the last or best one.

The almost strong correlation (P_{XY} 0.69) between the students' number of trials and the overall average assessment can be explained due to the idea that learners receive guidance in terms of run-time feedback messages whenever they make a wrong interaction with the game (i.e., violating any of the fire evacuation rules such as collecting personal belongings or using elevators). Therefore, in the next time they play the game, students try to avoid previous mistakes and thus achieve better results. Moreover, the assessment approach is also evaluating the performance of some skills such as crawling in areas full of smoke, such skills require a behaviour that players cannot keep steady every time they play the game and requires many times of repetition in order to gain the required skills, thus the low average overall assessment. However, it is worth noting that three schools out of four have strong correlation values ($P_{XY} > 0.7$). For instance, school 2 which has almost half of the experiment participants of (21 students out of 45), has the highest correlation value or (P_{XY} 0.76).

Regarding the suggestions coming out of two tutors, to improve the usability of the game as they noticed that on some older PCs the game had unsmooth interaction, unfortunately, addressing this issue is non-trivial and can not be done by simply reducing the games visual fidelity. Creating games with lower fidelity to meet the specifications of lab hardware risks to make the games themselves little compelling for users familiar with high-fidelity entertainment gaming. Therefore, a proper solution to this issue would be a reduction of the gap in hardware availability between school and leisure times. However, apart from this case of school's limited labs' hardware capabilities, and in contrast to findings from literature (Kickmeier-Rust and Albert 2010), the performance of the integrated and dynamic assessment and feedback within the serious game had no negative comments from participants or notes from the tutors. This is due to the service-oriented architecture

(SOA) based design (AL-Smadi and Guetl 2011; Dagger et al. 2007; Davies and Davis 2005). The performance of evaluating players' interactions and providing feedback was nearly similar to the game embedded assessment (based on game scripts).

6 Conclusion and future work

This research proposes an enhanced integrated dynamic assessment with serious games to support emergency training (i.e., fire evacuation). The proposed assessment approach consists of (1) an assessment model through which XML based representation of assessment rules are provided, the assessment rules use the game engine annotations of the pedagogical objects (e.g., bag, elevator, stairs, etc.) to design assessment based on possible player behaviour pattern and consequences (e.g., textual feedback, dialog with a non-player character (NPC), etc.), (2) an assessment engine that evaluates the event received from the game engine against the assessment rules defined for the game application domain (e.g., fire evacuation training) and provides feedback. The feedback is provided through a NPC within the game engine (the NPC may represent a teacher or a firefighter depending on the game scenario), (3) assessment interface handles the communication between the game engine and the assessment engine, and finally (4) a log file holds all tracked player interactions and environment changes, the log file is used for post evaluation to provide report based on player behaviour and performance within the game environment.

Experimentation results show the applicability of the proposed assessment approach in enhancing emergency training using serious games. Moreover, students were engaged to the proposed learning scenarios and their overall fire evacuation assessment were enhanced using the guided exploratory game-based training. The flexible design of the proposed assessment approach (i.e. SOA-based) could make it applicable to other interactive, event-based learning and training systems not necessarily relying upon three-dimensional gaming, even targeting professionals or domain-specific users and students. Among the possible evolutions, the authors are considering the application of the same methodologies in the domain of legal training as evolution of the systems described in (Arosio et al. 2013; Capuano et al. 2015).

A major challenge that may face educators and professional trainers using our approach is the authoring phase of the domain-specific assessment rules (represented by the assessment module). Representing/authoring assessment rules as part of the assessment module requires technical expertise as well as comprehensive understanding of the assessment model meta-data (XML-based rules and tags). Therefore, as future work we are planning to develop an assessment authoring module. This module will be an interactive graphical user interface (GUI) that can be easily integrated within the LMS. Moreover, this GUI will be flexible to interface with domain-specific knowledge bases (mainly using domain ontologies) to support the authoring of the assessment rules as discussed in Sect. 3.3 (Assessment model).

References

AL-Smadi M (2015) GAMEDUCATION: using gamification techniques to engage learners in online learning. Springer International Publishing, Cham

AL-Smadi M, Guetl C (2011) Service-oriented flexible and interoperable assessment: towards a standardised e-assessment system. Int J Contin Eng Educ Life Long Lear 21(4):289–307

AL-Smadi M, Wesiak G, Guetl C (2012) Assessment in serious games: an enhanced approach for integrated assessment forms and feedback to support guided learning. In: 15th international conference on interactive collaborative learning (ICL), 2012. IEEE, pp 1–6

Almond R, Steinberg L, Mislevy R (2002) Enhancing the design and delivery of assessment systems: a four-process architecture. J Technol Learn Assess 1(5)

Andrée K, Nilsson D, Eriksson J (2015) Evacuation experiments in a virtual reality high-rise building: exit choice and waiting time for evacuation elevators. Fire Mater

Arosio G, Bagnara G, Capuano N, Fersini E, Toti D (2013) Ontology-driven data acquisition: intelligent support to legal ODR systems. In: JURIX, pp 25–28

Averill JD, Mileti DS, Peacock RD, Kuligowski ED, Groner N, Proulx G, Reneke PA, Nelson HE (2005) Federal building and fire safety investigation of the world trade center disaster occupant behavior, egress, and emergency communications. Tech. rep., National Institute of Standards and Technology, Washington, DC

Bloom BS, Engelhart MD (1969) Taxonomy of educational objectives: the classification of educational goals: by a Committee of College and University Examiners: handbook 1. David McKay

Bloomfield PR, Livingstone D (2009) Immersive learning and assessment with quizhud. Comput Inf Syst J 13(1):20–26

Brooke J et al (1996) Sus-a quick and dirty usability scale. Usability Eval Ind 189(194):4–7

Bukowski RW (2012) Addressing the needs of people using elevators for emergency evacuation. Fire Technol 48(1):127–136

Bunchball I (2010) Gamification 101: an introduction to the use of game dynamics to influence behavior. White paper

Burgos D, Moreno-Ger P, Sierra JL, Fernández-Manjón B, Specht M, Koper R (2008) Building adaptive game-based learning resources: The integration of IMS learning design and e-adventure. Simulation and Gaming

Capuano N, King R (2015) Knowledge-based assessment in serious games: an experience on emergency training. J e-Learn Knowl Soc 11(3)

Capuano N, Longhi A, Salerno S, Toti D (2015) Ontology-driven generation of training paths in the legal domain. Int J of Emerging Technol Learn 10(7):14–22

Chang M, et al. (2010) Web-based multiplayer online role playing game (morpg) for assessing students' java programming knowledge and skills. In: Digital game and intelligent toy enhanced learning (DIGITEL), 2010 third IEEE international conference on, IEEE, pp 103–107

Chittaro L, Ranon R (2009) Serious games for training occupants of a building in personal fire safety skills. In: Games and virtual worlds for serious applications, 2009. VS-GAMES'09. Conference in, IEEE, pp 76–83

Connolly TM, Stansfield M, Hainey T (2011) An alternate reality game for language learning: arguing for multilingual motivation. Comput Educ 57(1):1389–1415

Dagger D, O'Connor A, Lawless S, Walsh E, Wade VP (2007) Service-oriented e-learning platforms: from monolithic systems to flexible services. IEEE Internet Comput 11(3):28–35

Davies WM, Davis HC (2005) Designing assessment tools in a service oriented architecture. In: 1st international ELeGI conference on advanced technology for enhanced learning, p 7

De Freitas S, Neumann T (2009) The use of exploratory learningfor supporting immersive learning in virtual environments. Comput Educ 52(2):343–352

De Freitas S, Oliver M (2005) A four dimensional framework for the evaluation and assessment of educational games. Comput Assist Learn

De Jong T, Van Joolingen WR (1998) Scientific discovery learning with computer simulations of conceptual domains. Rev Educ Res 68(2):179–201

Deterding S, Dixon D, Khaled R, Nacke L (2011) From game design elements to gamefulness: defining gamification. In: Proceedings of the 15th international academic MindTrek conference: envisioning future media environments, ACM, pp 9–15

Ding Y, Yang L (2013) Occupant evacuation process study of public buildings based on computer modeling and simulation. J Appl Fire Sci 23(3):365–380

Ding Y, Yang L, Weng F, Fu Z, Rao P (2015) Investigation of combined stairs elevators evacuation strategies for high rise buildings based on simulation. Simul Model Pract Theory 53:60–73

Dunwell I, de Freitas S, Jarvis S (2011) Four-dimensional consideration of feedback in serious games. In: de Freitas S, Maharg P (eds) Digital games and learning. Continuum Publishing, pp 42–62

Eseryel D, Ifenthaler D, Ge X (2011) Alternative assessment strategies for complex problem solving in game-based learning environments. In: Ifenthaler D, Michael Spector J, Isaias P, Sampson D, Kinshuk (eds) Multiple perspectives on problem solving and learning in the digital age. Springer, New York, pp 159–178

Fahy RF (1994) Exit 89-an evacuation model for high-rise buildings-model description and example applications. Fire Saf Sci 4:657–668

Fahy RF, Proulx G (2009) 'Panic' and human behaviour in fire. pp 387–398

Fosnot CT, Perry RS (1996) Constructivism: a psychological theory of learning. In: Fosnot CT (ed) Constructivism: theory, perspectives, and practice. Teachers College Press, New York, pp 8–33

Francis R, Saunders PB (1979) EVACNET: prototype network optimization models for building evacuation. National Bureau of Standards, Washington, DC

Friedler Y, Nachmias R, Linn MC (1990) Learning scientific reasoning skills in microcomputer-based laboratories. J Res Sci Teach 27(2):173–192

Girard C, Ecalle J, Magnan A (2013) Serious games as new educational tools: how effective are they? A meta-analysis of recent studies. J Comput Assist Learn 29(3):207–219

Glaser R, Schauble L, Raghavan K, Zeitz C (1992) Scientific reasoning across different domains. In: De Corte E, Linn M, Mandl H, Verschaffel L (eds) Computer-based learning environments and problem solving. Springer, Berlin, pp 345–373

Gwynne S, Galea E, Owen M, Lawrence PJ, Filippidis L (1999) A review of the methodologies used in the computer simulation of evacuation from the built environment. Build Environ 34(6):741–749

Hainey T, Connolly TM, Stansfield M, Boyle EA (2011) Evaluation of a game to teach requirements collection and analysis in software engineering at tertiary education level. Comput Educ 56(1):21–35

Hainey T, Connolly TM, Chaudy Y, Boyle E, Beeby R, Soflano M (2014) Assessment integration in serious games. In: Connolly TM, Hainey T, Boyle E, Baxter G, Moreno-Ger P (eds) Psychology, pedagogy, and assessment in serious games. IGI Global, Hershey, pp 317–341. doi:10.4018/978-1-4666-4773-2.ch015

Hunicke R, LeBlanc M, Zubek R (2004) Mda: A formal approach to game design and game research. In: Proceedings of the AAAI workshop on challenges in game AI, vol. 4, p 1

Ibáñez MB, Crespo RM, Kloos CD (2010) Assessment of knowledge and competencies in 3D virtual worlds: a proposal. In: Reynolds N, Turcsányi-Szabó M (eds) Key competencies in the knowledge society. Springer, Berlin, pp 165–176

Kemp JW, Livingstone D, Bloomfield PR (2009) Sloodle: connecting vle tools with emergent teaching practice in second life. Brh J Educ Technol 40(3):551–555

Kendik E (1986) Methods of design for means of egress: towards a quantitative comparison of national code requirements. Fire Saf Sci 1:497–511

Ketchell N, Cole S, Webber D, Marriott C, Stephens P, Brearley I, Fraser J, Dohney J, Smart J (1993) The EGRESS code for human movement and behaviour in emergency evacuations. University of Edinburgh, Artificial Intelligence Applications Institute

Kickmeier-Rust MD, Albert D (2010) Micro-adaptivity: protecting immersion in didactically adaptive digital educational games. J Comput Assist Learn 26(2):95–105

Kickmeier-Rust MD, Steiner CM, Albert D (2009) Non-invasive assessment and adaptive interventions in learning games. In: Intelligent networking and collaborative systems, 2009. INCOS'09. International conference on, IEEE, pp 301–305

Kisko TM, Francis R, Nobel C (1998) Evacnet4 users guide. University of Florida, Gainesville

Kobes M, Helsloot I, De Vries B, Post JG (2010) Building safety and human behaviour in fire: a literature review. Fire Saf J 45(1):1–11

Kolb D (1984) Experiential learning: experience as the source of learning and development. Prentice Hall, Englewood Cliffs

Korhonen T, Hostikka S (2009) Fire dynamics simulator with evacuation: Fds+ evac. Technical reference and users guide. VTT Technical Research Centre of Finland

Kuligowski ED, Gwynne SM (2010) The need for behavioral theory in evacuation modeling. In: Klingsch WWF, Rogsch C, Schadschneider A, Schreckenberg M (eds) Pedestrian and evacuation dynamics 2008. Springer, Berlin, pp 721–732

Lave J, Wenger E (1991) Situated learning: legitimate peripheral participation. Cambridge University Press, New York

MacDonald M (2003) Steps simulation of transient evacuation and pedestrian movements. User manual, Unpublished Work

Maderer J, Gütl C, Al-Smadi M (2013) Formative assessment in immersive environments: A semantic approach to automated evaluation of user behavior in open wonderland. J Immers Educ–Proceedings of iED 2013 Boston Summit, Boston

McAlpine M, van der Zanden L, Harris V, Authority SQ (2010) Using games based technology in formal assessment of learning. In: Proceedings of the 4th European conference on games-based learning: ECGBL2010, Academic Conferences Limited, p 242

Mislevy R, Riconscente M (2005) Evidence-centered assessment design: Layers, structures, and terminology (PADI technical report 9)

Mislevy RJ, Almond RG, Lukas JF (2004) A brief introduction to evidence-centered design. CSE report 632. US Department of Education

Moreno-Ger P, Burgos D, Martínez-Ortiz I, Sierra JL, Fernández-Manjón B (2008) Educational game design for online education. Comput Human Behav 24(6):2530–2540
Mory E (2006) Feedback research revisited. In: Jonassen DH (ed) Handbook of research for educational communications and technology. Macmillan, New York, pp 745–783
Njoo M, De Jong T (1993) Exploratory learning with a computer simulation for control theory: learning processes and instructional support. J Res sci Teach 30(8):821–844
Raybourn EM (2007) Applying simulation experience design methods to creating serious game-based adaptive training systems. Interact Comput 19(2):206–214
Ribeiro J, Almeida JE, Rossetti RJ, Coelho A, Coelho AL (2012) Using serious games to train evacuation behaviour. In: Information systems and technologies (CISTI), 2012 7th Iberian Conference on, IEEE, pp 1–6
Rogers C (1951) Client-centered therapy: its current practice, implications and theory. Constable, London
Ronchi E, Nilsson D (2013) Fire evacuation in high-rise buildings: a review of human behaviour and modelling research. Fire Sci Rev 2(1):1–21
Santos G, Aguirre BE (2004) A critical review of emergency evacuation simulation models. Tech. rep., National Institute of Standards and Technology, Washington, DC
Shute VJ (2008) Focus on formative feedback. Rev Educ Res 78(1):153–189
Shute VJ, Ventura M, Bauer M, Zapata-Rivera D (2009) Melding the power of serious games and embedded assessment to monitor and foster learning. Serious Games Mech Eff 2:295–321
Sliney A, Murphy D (2008) Jdoc: a serious game for medical learning. In: Advances in computer–human interaction, 2008 first international conference on, IEEE, pp 131–136
Sliney A, Murphy D (2011) Using serious games for assessment. In: Ma M, Oikonomou A, Jain LC (eds) Serious games and edutainment applications. Springer, London, pp 225–243
Sliney A, Murphy D, O'Mullane J (2009) Secondary assessment data within serious games. In: Petrovic O, Brand A (eds) Serious games on the move. Springer, Vienna, pp 225–233
Swaak J, Van Joolingen WR, De Jong T (1998) Supporting simulation-based learning; the effects of model progression and assignments on definitional and intuitive knowledge. Learn Instr 8(3):235–252
Tan L, Hu M, Lin H (2015) Agent-based simulation of building evacuation: combining human behavior with predictable spatial accessibility in a fire emergency. Info Sci 295:53–66
Underwood JS, Kruse S, Jakl P (2010) Moving to the next level: designing embedded assessments. In: Zemliansky P, Wilcox D (eds) Design and implementation of educational games: theoretical and practical perspectives. Information Science Network, Hersey, New York, pp 126–140
Veenman MV, Elshout JJ, Meijer J (1997) The generality vs domain-specificity of metacognitive skills in novice learning across domains. Learn Instr 7(2):187–209
Vygotsky LS (1986) Thought and language (revised edition, Kozulin A (ed)). MIT Press, Cambridge, MA
Williams-Bell F, Kapralos B, Hogue A, Murphy B, Weckman E (2015) Using serious games and virtual simulation for training in the fire service: a review. Fire Technol 51(3):553–584
Zhang W, Huang L, Wang B (2009) Application of the EVACNET4 model to evacuation in high-rise building. Fire Sci Technol 3:014