Serious Games - Evaluation

Thomas Hainey

Prof. Thomas Connolly
School of Computing
University of West of Scotland

Thomas.Connolly@UWS.ac.uk
Thomas.Hainey@UWS.ac.uk

Structure of Presentation

- Evaluation
  - Requirement for evaluation of Serious Games
  - Why evaluate Serious Games?
  - Formative Evaluation
  - Summative Evaluation
- Important Questions to ask when evaluating Serious Games
- Existing evaluation frameworks
- Compiled evaluation framework

Evaluation

- Gavin (2008) defines evaluation as: *"Research involving systematic appraisal of organisations, processes or programmes leading to feedback on improvement or performance."*

- Evaluation of products or Serious Game interventions can also occur

- Two main types of evaluation: formative and summative

Requirement for evaluation of Serious Games

- Serious Games have captured the interest of educationalists and industrialists
- There is a dearth of empirical evidence supporting the approach
- A lack of general frameworks from a pedagogical perspective
- Existing frameworks do not specify particular measurements that can be taken as a starting point for researchers

Why evaluate Serious Games?

- Improve usability of Serious Games
- Inform theory
- Increase user acceptance
- To assist in selling your Serious Game
- Enhance learning outcomes
- Increase learning efficiency
- To assist in the development of the Serious Games field

Evaluation - Formative Evaluation

- Ogle (2002) states that formative evaluation is "*a systematic and empirical process although rarely a scientific one*"

- One advantage of formative evaluation is that it also allows evaluation of the instruments of evaluation.

Evaluation - Formative Evaluation

- *Expert Reviews* – Conducted very early on in the evaluation process and is particularly focused on instructional content, technical quality or accuracy.

- *One-To-One Evaluations* – Designed for the developer to work with a number of potential learners from the intended user base, primarily to assess the learners' reaction to content and assess particular indicators of performance.

Evaluation - Formative Evaluation

- *Small Group Evaluations* – Conducted using small groups in which the instructor interacts with the learners in the same type of environmental context as the intervention will be used.

- *Field trials* – Also known as field-tests or beta tests, field trials are designed to see if the changes made from the small group evaluation were effective and whether the intervention can be used in the intended context.

Evaluation - Summative Evaluation

- Summative evaluation is to assess an end product and as a result this type of evaluation is usually performed by external evaluators.

- Research-oriented summative evaluations are to validate and improve programs.

- Management-oriented summative evaluations are to assess cost and whether the programs did what they were supposed to do.

Important Questions to ask when evaluating Serious Games

- What is to be done with the information collected?
- What are the appropriate forms of measurement?
- What is the most appropriate experimental design?
- What is an appropriate form of comparison?
- What is an appropriate context?

# Evaluation –What is to be done with the information collected?

- To assess if the learners' learn the subject matter better by playing a Serious Game
- To assess the usability of the Serious Game
- To assess attitudinal or behavioural change as a result of playing a Serious Game
- To assess likeability, acceptability etc

Evaluation – What are the appropriate forms of measurement?

- Measurements may include:
  - Increase in knowledge –pre-test/post-test
  - Efficiency of task completion
  - What aspects of the Serious Game are utilised?
  - Attitudes to the Serious Game
  - Savings with regards to time and cost
- Generally measurements are acquired through questionnaires, interviews observation, log files etc

Evaluation – What is the most appropriate experimental design?

- Experimental or quasi-experimental?
- Pre-test (possibly to determine if the population sample is adequate)
- Intervention → post-test
- Pre-test → Intervention → post-test
- Pre-test → Intervention → post-test → long term follow up post-test
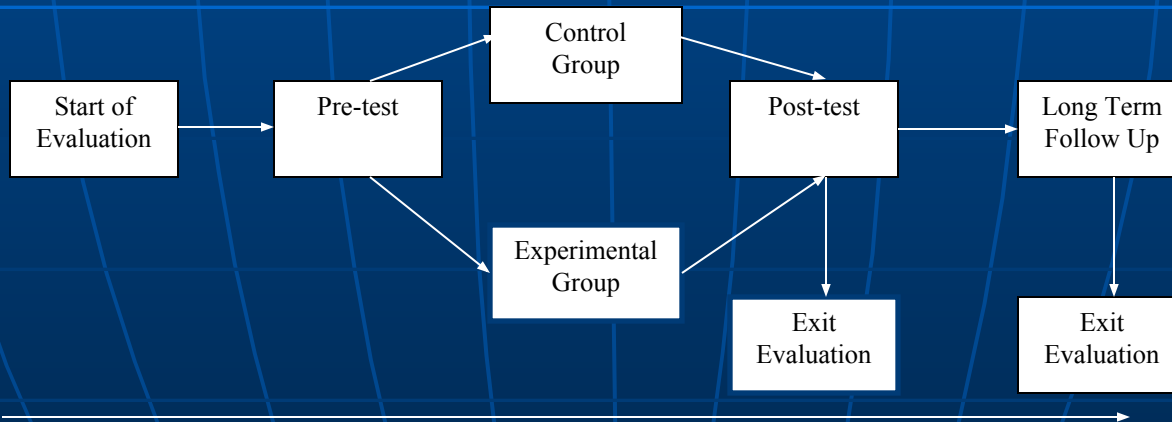
Evaluation – What is an appropriate form of comparison?

- Serious Game on it's own
- Serious Game vs. Traditional Teaching Techniques
  - Text books, paper-based case studies
  - Live-through case studies
  - Role-playing
  - Lectures
  - Labs
  - It is more beneficial to utilise a control group

Evaluation – What is an appropriate context?

- Labs
- Tutorials
- Lectures
- Home
- The context where the Serious Game is designed to be used is generally most effective
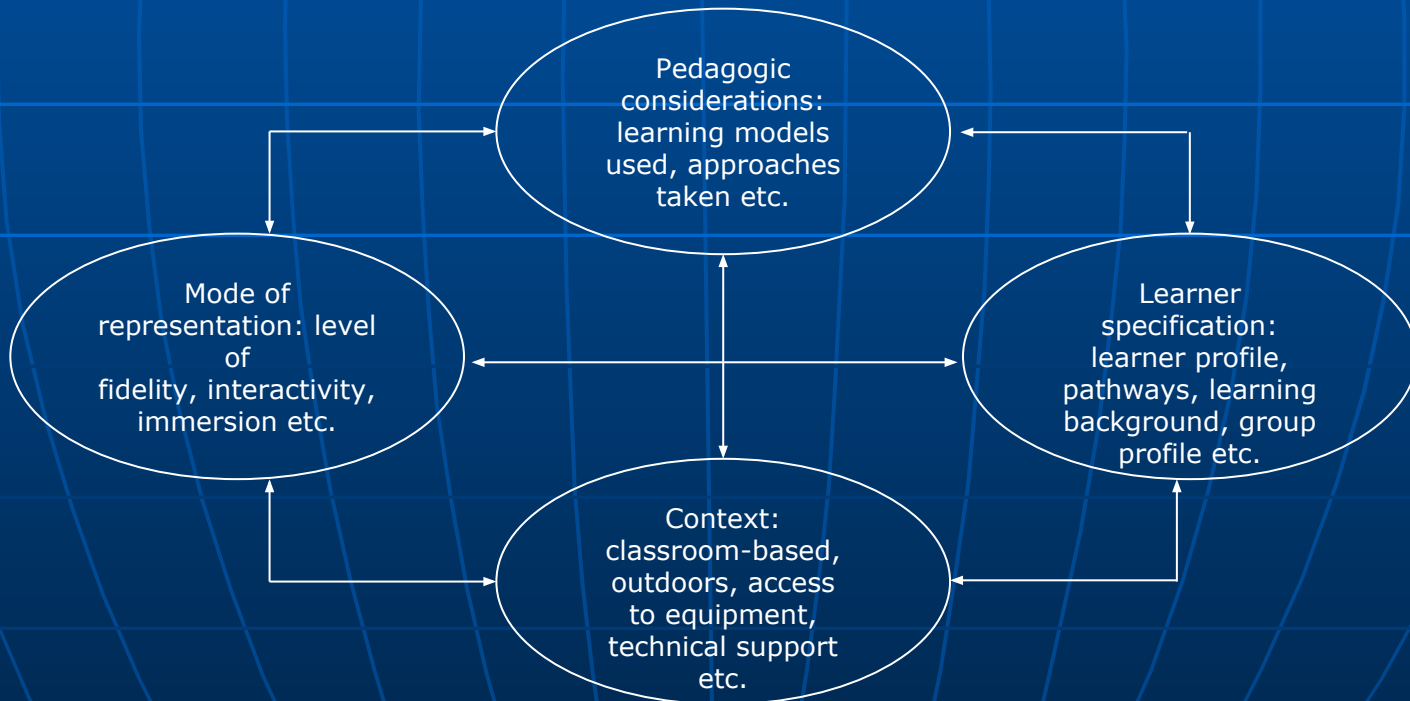
Previous evaluation frameworks

- Frameworks based on Heuristic Evaluation
  - Web-based specific criteria
  - MMORPGs
- Problems associated are:
  - Heuristic evaluation is based on expert reviewers knowledge
  - Primarily focused on usability and interface problems

Previous evaluation frameworks

- Tan *et al* (2007) reviewed four GBL frameworks
  - Design framework for educational environments
  - Adopted interaction cycle for games
  - Engaging multimedia design model for children
  - Game Object Model
- Only the Game Object Model significantly addressed pedagogy
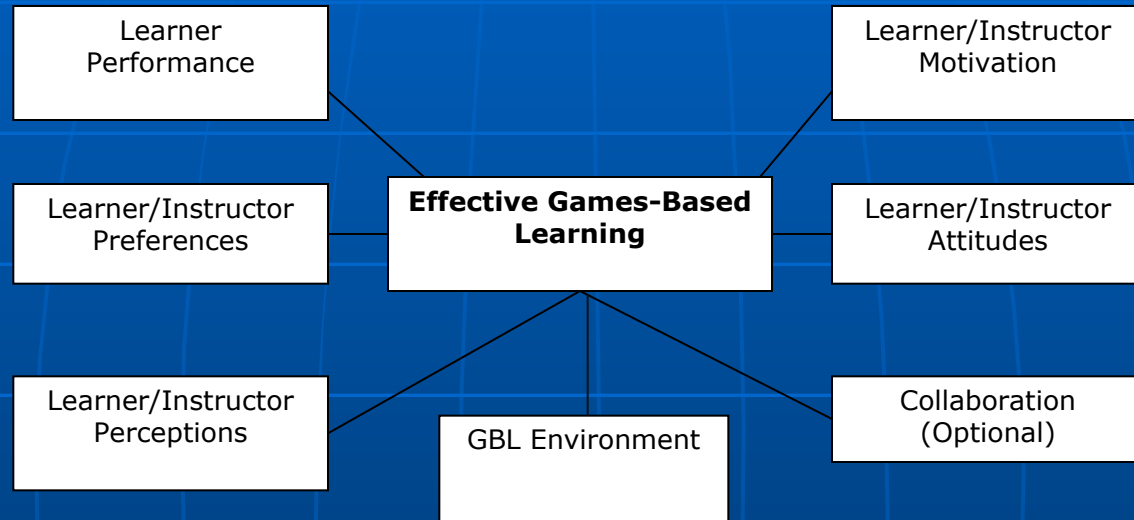
Previous evaluation frameworks

Level 1: REACTION	Trainee's reaction to the program: level of satisfaction
Level 2: LEARNING	Trainee's attitude change, increased knowledge, and/or increased skill, due to the training
Level 3: BEHAVIOUR	On the job change in behaviour because of program participation, i.e. transfer of learning to the job setting
Level 4: RESULTS	How the organization benefited from the learner's participation in the program (e.g. increased profits)

Previous evaluation frameworks

Compiled evaluation framework for effective Serious Game Evaluation

Learner Performance	Learner/Instructor Motivation
Learner/Instructor Preferences	**Effective Games-Based Learning**	Learner/Instructor Attitudes
Learner/Instructor Perceptions	GBL Environment	Collaboration (Optional)

- Learner Performance – encompasses pedagogy from a learner perspective. It is primarily concerned with whether there is an improvement in learner performance
- Learner/Instructor Motivation – particular motivations of the learner and instructor for using the intervention

Compiled evaluation framework for effective GBL

- Learner/Instructor Preferences – encompasses different learner and instructor preferences (Kolb 1984)
- Learner/Instructor Attitudes – various elements that may alter the effectiveness of the GBL intervention
- Learner/Instructor Perceptions – encompasses how the learner and instructor perceives the intervention
- Collaboration – optional as it is dictated by how the game is played

Compiled evaluation framework for effective GBL

- GBL environment category is split into: environment, scaffolding, usability, level of social presence and deployment

  - Environment – background environment, characters, environmental alteration

  - Scaffolding – advice quality, resource allocation

  - Usability – task completion times, ease of use

Compiled evaluation framework for effective Serious Game Evaluation

- Level of social presence – immersion and interaction in the game world
- Deployment – effective method of appropriate incorporation, preferences of different gaming conditions

Measurements associated with Learner Performance

- Acquisition of procedural, declarative, explicit, implicit or general knowledge
- Formation of metacognitive strategies
- Skill acquisition
- Characteristics of participants at baseline
- Measurement of the advantages and the disadvantages of using the Serious Game

Measurements associated with Learner/Instructor Perceptions

- Learners perception of how a Serious Game represents a holistic view of an organisation or process

- Perception of complexity, fluency, feedback and realism

- Instructors perception of how well the Serious Game fits into their course

- Instructors perception of how the Serious Game saves time

Measurements associated with Learner/Instructor Preferences

- Learner preferences of learner styles: competitive, cooperative, individualistic
- Rating of technical aspects of the Serious Game
- Learner/Instructor likes and dislikes of conventional training
- Preferred activities within the environment

Measurements associated with Learner/Instructor Motivation

- Measurement of willingness to play over a prolonged period of time
- Level of interest in playing initially
- Features that make the Serious Game more interesting
- Instructors motivation for using a Serious Game in their course

Measurements associated with Learner/Instructor Attitudes

- Learners/Instructors attitude towards Serious Games
- Learners attitude towards the subject being taught
- Learner/Instructors attitudes towards game elements such as interface, sound usefulness of feedback etc
- Attitudes, motivations and preferences could be considered if the Serious Game fails to act as a learning tool

Measurements encountered associated with GBL Environment

- **<u>Environment</u>**
- Background environment validation
- Virtual environment navigation
- Character validation
- Assessing and gauging virtual agent expressiveness
- Assessing the context in terms of decision-making support as found in the real world
- Advice importance
- Environment usability, acceptability, credibility

Measurements encountered associated with GBL Environment

- **<u>Scaffolding</u>**
- Average amount of times scaffolding/advice resources were used
- Expert review knowledge of advice quality
- Students perceptions of advice quality
- Appropriate feedback and realism

Measurements encountered associated with GBL Environment

- **<u>Usability</u>**
- User performance
- User confidence
- Player reactions to first and incremental prototypes
- Average task completion time, ease of tasks, ease of use

Measurements encountered associated with a GBL environment

- **<u>Level of social presence</u>**
- Frequency of relationships
- Learner evaluation of game character's personalities
- Mood and attitude statements towards characters and events
- Level of immersion and interaction
- **<u>Deployment</u>**
- Preference of different gaming conditions
- Method of appropriate incorporation into an educational context

Measurements associated with Collaboration

- Frequency/Level of collaboration
- Collaboration in different groups such as competitive and cooperative
- Mapping of particular comments to team aspects – dialogue analysis