GaLA
Game and Learning Alliance
The European Network of Excellence on Serious Games

Workshop
Are you serious?
Evidence for Learning using Games

Rob Nadolski, Open University of the Netherlands
Jannicke Baalsrud Hauge, BIBA, Bremer Institut für Produktion und Logistik
Elizabeth Boyle, University of the West of Scotland
Johann Riedel, The University of Nottingham
Igor Mayer, Delft University of Technology
Pablo Moreno-Ger, Universidad Complutense de Madrid
Workshop Agenda

#OEB12  #WSSGE

• 0. Welcome & Intro Assessment WITH(IN) Serious Games (Rob)
   10:00- 10:10

• 1. Evaluation of Serious Games (Liz)
   10:10-10:25

• 2. Challenges for assessment within Serious Games (Rob)
   10:25-10:35

• 3. Best practices & guidelines using games for learning (all)
   10:35-11:00

• 4. Hands-on session evaluating games (all)
   11:00- 12:45 (break at 11:30)

• 5. Report Hands-on & Discussion (Jannicke)
   12:45-13:00
GaLA
Game and Learning Alliance
The European Network of Excellence on Serious Games

Intro Assessment WITH(IN) Serious Games

Rob Nadolski
Open University of the Netherlands
Centre for Learning Sciences and Technology
According to Corti (2011):

“Serious games will only grow as an industry if the learning experience is definable, quantifiable and measurable.”
What is assessment and its main purpose?

Process of using data to *demonstrate* that stated (learning) goals and objectives are actually being met by a person(learner).

(Chin, Dukes & Gamson, 2009)

Assessments need to be aligned with learning objectives

(Gagné, Briggs, & Wager, 1988; Smith & Ragan, 1999)
Types of assessment in learning

• Formative assessment (FA)  during learning (process)
  Purpose: gather evidence for actions to improve learning
  = assessment FOR learning

• Summative assessment (SA)  after learning (product)
  Purpose: gather evidence of learning
  = assessment OF learning

• FA & SA differ in purpose but may use similar methods
Types of assessment in GBL- Serious games (SGs)

(Ifenthaler, Eseryel, & Ge, 2012)

• **External assessment** (external to the game: disruptive) *before, during, and after learning*
  - (de) briefing interviews
  - knowledge maps or causal diagrams
  - multiple-choice questions or essays

• **Embedded assessment** (in the game: should be non-disruptive) *during learning and gaming*
  - clickstreams, logfiles
  - information trials

• **Game scoring** (in the game: non-disruptive) *during and after gaming*
  - targets achieved or obstacles overcome while playing
  - time needed for completing a specific task
Issues with types of assessment in GBL-SGs 

(Ifenthaler, Eseryel, & Ge, 2012)

1. External assessment [external to the game] overall(-)
   - inefficient: in-game data remain unused (--)
   - inefficient: scores do not affect the game (--) 
   - disruptive for the experience (flow, engagement) (--) 
   - only product - no process → incomplete evidence of learning & not scalable (grade-focused instead of learning focused)(--)

   [summative assessment of learning]

2. Embedded assessment [in the game] overall(+)
   - non-disruptive adaptive gaming if design is balanced (+++)
   - reliability and validity → challenges technology & methods(+-)
   - increases development costs, but is scalable (-)

   [formative & summative assessment for/of learning and gaming]

3. Game scoring [in the game] overall(-)
   - non-disruptive, but often does NOT measure learning (-----)
Motives for Embedded Assessment in GBL-SGs

1. Gives insight into underlying learning processes when playing that guides offering scaffolds to improve learning
2. Enables better understanding of specific behavior and the final outcomes (tracking motivation, emotions, a.s.o.)
3. Can point to specific areas of difficulties learners experience
4. Points out the strengths and the weaknesses of the game

Embedded assessment in SGs
- needs dedicated design (balance: play, learn, meaning).
- can offer challenges for technology & methods (reliable, valid)

WHAT, HOW and CONTEXT of assessment DO matter.
GaLA
Game and Learning Alliance
*The European Network of Excellence on Serious Games*

**Evaluation of serious games**

*Dr. Elizabeth Boyle*
*Prof Thomas Connolly*
*Dr. Thomas Hainey*
*Prof James Boyle*
Assessment and evaluation
(Gikandi, Morrow and Davis, 2011)

both assessment and evaluation have a component of measurement but:

assessment refers to:
“… operations associated with measuring achievements of persons in relation to desirable outcomes”.

evaluation refers to:
“… operations associated with measuring worthiness /value of non-person entities (such as curricula, programmes, courses, instructional strategies among others) in relation to identified goals”
or
“Research involving systematic appraisal of organisations, processes or programmes leading to feedback on improvement or performance”.
(Gavin, 2008)
Evaluation of educational interventions proposes a hierarchy of evidence (Woolfson, 2011)

- Meta-analyses of data from systematic reviews
- Randomised controlled trials (RCT)
- Quasi-experimental designs
- Single case experimental designs – pre post test
- Non experimental designs – surveys, correlational, qualitative
RCT: Design of choice?

- In evaluating the effectiveness of a game for learning or behaviour change participants are randomly allocated to an experimental [E] or control [C] condition and the target skill/behaviour is measured for both groups before and after participation in the intervention (game).
  - Pre-testing should confirm no existing difference between E and C.
  - Post-testing should show that E performs better than C.
- Improvements in target skill/behaviour for E compared with C in a follow-up study would allow further confirmation that the intervention was successful.
- But a trial should be ‘ethically neutral’: we should not know in advance whether the intervention is helpful.
Problems for RCT

Problems

• Frequently difficult to randomly allocate participants
• Randomization often at class level with the effective sample size the number of classes randomized (c.f. cluster RCT)
• Ethical issues: if the game works in a classroom setting, game group will be advantaged in assessment! Though not a problem if we do not know in advance whether the game will be helpful

One solution

• All players play a version of the game with some tackling topic a and others tackling topic b; all players get some advantage of playing the game. Compare performance of groups on topics a and b on the different issues
Aims:

• to determine what **empirical** research has been carried out on outcomes and impacts of digital games (entertainment games, games-based learning and serious games)

• to examine what outcomes and impacts of playing games have been examined

• to determine how best to organise and categorise the diverse research in this area

• to tackle defragmentation
Stages in systematic literature review

Search terms
(“computer games” OR “video games” OR “serious games” OR “simulation games” OR “games-based learning” OR MMOG OR MMORPG OR MUD OR “online games”)
AND (evaluation OR impacts OR outcomes OR effects OR learning OR education OR skills OR behaviour OR attitude OR engagement OR motivation OR affect)

Time period: January 2004 to February 2009

Databases/journals searched
ACM, ASSIA, BioMed Central, Cambridge Journals Online, ChildData, Index to Theses, Oxford University Press (journals), ScienceDirect, EBSCO, SocINDEX, Library, Information Science and Technology Abstracts, CINAHL), ERIC, IngentaConnect, Infotrac, Emerald and IEEE

Selection of papers about empirical evidence
Quality rating
Coding of selected papers
### Number of papers providing empirical evidence for each learning and behavioural outcome

<table>
<thead>
<tr>
<th>Outcomes of playing games</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affective and motivational outcomes</td>
<td>18</td>
</tr>
<tr>
<td>Knowledge acquisition/content understanding</td>
<td>17</td>
</tr>
<tr>
<td>Perceptual and cognitive skills</td>
<td>13</td>
</tr>
<tr>
<td>Behaviour change</td>
<td>8</td>
</tr>
<tr>
<td>Physiological outcomes</td>
<td>6</td>
</tr>
<tr>
<td>Social/soft skill outcomes</td>
<td>4</td>
</tr>
<tr>
<td>Motor skills</td>
<td>4</td>
</tr>
<tr>
<td><strong>Grand Total (out of 7,392 papers)</strong></td>
<td><strong>70</strong></td>
</tr>
</tbody>
</table>
## Primary purpose of game by learning and behavioural outcome

<table>
<thead>
<tr>
<th>Outcomes of playing game</th>
<th>Entertainment game</th>
<th>Game for learning</th>
<th>Serious game</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affective and motivational outcomes</td>
<td>14</td>
<td>4</td>
<td>0</td>
<td>18</td>
</tr>
<tr>
<td>Knowledge acquisition/content understanding</td>
<td>3</td>
<td>12</td>
<td>2</td>
<td>17</td>
</tr>
<tr>
<td>Perceptual and cognitive skills</td>
<td>9</td>
<td>4</td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td>Behaviour change</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>Physiological outcomes</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Social/soft skill outcomes</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Motor skills</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td><strong>Grand Total</strong></td>
<td><strong>36</strong></td>
<td><strong>26</strong></td>
<td><strong>8</strong></td>
<td><strong>70</strong></td>
</tr>
</tbody>
</table>
## Primary purpose of game by study design

<table>
<thead>
<tr>
<th>Study Design</th>
<th>Purpose of game</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Entertainment</td>
<td></td>
</tr>
<tr>
<td>Quasi-experimental</td>
<td>16</td>
<td>34</td>
</tr>
<tr>
<td>Survey</td>
<td>16</td>
<td>22</td>
</tr>
<tr>
<td>RCT</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>Qualitative Study</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>36</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>Game for Learning</td>
<td></td>
</tr>
<tr>
<td></td>
<td>14</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>26</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>70</td>
<td></td>
</tr>
</tbody>
</table>
Process models: evaluation before, during and after play

Before
Performance on relevant knowledge/skill
Motives: self determination theory; uses and gratifications
Technology acceptance model: perceived usefulness and ease of use of technology
Theory of planned behaviour: players’ attitudes to games, subjective norms and perceived control

During
Performance on relevant knowledge/skill
Models of user experience: provide a broader picture of how players feel while playing games including flow, presence, immersion, arousal
Models of usability: effectiveness, efficiency and user satisfaction

After
Performance on relevant knowledge/skill
Transfer of knowledge and skills to the real world
Impact of game on organisation
Issues in evaluation

• Evaluation of games is complex: twin goals of engagement and performance; if players are engaged is that sufficient?
• Formative or summative evaluation?
• Qualitative (subjective), quantitative (objective) or mixed methods?
• Must we rely on RCTs? Cluster RCT, where random allocation is at level of class or school may be appropriate in education (Tymms, Merrell and Coe, 2008)
• Quasi-experimental and qualitative designs can provide useful evidence
• Objective data are valued more but qualitative data can complement, support and confirm quantitative data
• Hierarchy of evidence for qualitative research too (Daly et al, 2007)
• Eclectic approach to establishing evidence base for classroom practice (Issett and Kyriacou, 2008)
• Difficult to identify a control for a game that offers new experience
GaLA
Game and Learning Alliance
The European Network of Excellence on Serious Games

Challenges for Assessment within Serious Games

Rob Nadolski
Open University of the Netherlands
Centre for Learning Sciences and Technology

Pablo Moreno Ger
Universidad Complutense de Madrid
Solutions for Embedded Assessment in GBL-SGs

• (1) The *Evidence Centered Design framework* (ECD) (Mislevy, et al. 2003; Shute et al., 2009)

• (2) Guidelines for well-designed assessment in games (Underwood, Kruse, & Jakl, 2010)

• (3) R&D and tooling for applying ECD in GBL (work in progress) (Shute et al, 2009; Shute & Kee, 2012; Westera et al.)

• (4) *Learning Analytics* (work in progress) [as opposite paradigm to top down approach(1,2,3)]

  “The measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs” (The Society for Learning Analytics Research)

• (5) Other…
  - FourDimensional Framework (de Freitas et al., 2010)
  - OKEI Competence modelling (TARGET-FP7 project, 2012)
(1) The **Evidence Centered Design framework** (ECD)

- Competency model (student’s knowledge, skills, attitudes, etc.)
- Evidence model (to inform inferences about the levels or states of competency model variables)
- Task model or Action model (activities/conditions: collecting data)
(1) The *Evidence Centered Design framework* (ECD)

**Advantage:**
- can cover all learning objectives
- is proven via several applications

**Disadvantage:**
- time-consuming for higher level skills
- applications for ‘cognitive states’ → lacking: affective states

**Issue:**
- choosing the right Granularity level
(2) Guidelines for well-designed assessment in games

(Underwood, Kruse, & Jakl, 2010)

- Provide custom feedback to learners during the game
- Provide custom feedback to learners after the game
- Tailor how learners experience the games (adaptive game play)

But….custom feedback needs feedback design

Feedback design decides on:

- type, content, format, & frequency
(3) R&D and tooling for applying ECD in GBL (example: sensors)

Framework for Embedded Assessment

(Westera & Nadolski work in progress)
OnLineEduca Berlin, November 28, 2012

European Commission Information Society and Media

Solutions for Embedded Assessment in GBL-SGs

(3) R&D and tooling for applying ECD in GBL (example: sensors)

Framework for Embedded Assessment

(Westera & Nadolski work in progress)
Solutions for Embedded Assessment in GBL-SGs

(4) Learning Analytics – definition & Framework
“The measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs”
(The Society for Learning Analytics Research, http://www.solaresearch.org/mission/about/)

---

Siemens (2011)
http://www.slideshare.net/fullscreen/gsiemens/learning-analytics-higher-education/1
(4) Learning Analytics (LA) – Techniques for Big Data problems

(Siemens & Gasevic, 2012)

- Information retrieval technologies, like:
  - educational data mining (EDM; cf. Romero et al., 2008)
  - machine learning
  - classical statistical analysis techniques
- Social Network Analysis (cf. Buckingham & Ferguson, 2011)
- Natural Language Processing (NLP)

Current applications of LA are restricted to logged student data within LMS but …. LA can be applied in GBL
Solutions for Embedded Assessment in GBL-SGs

(4) Learning Analytics (LA) – Framework in GBL – GLEANER
(Serrano et al., 2012)
(4) Learning Analytics - Affordances

• Supports learners, teachers & institutions

“Learning analytics need not simply focus on student performance. It might be used as well to assess curricula, programs, and institutions. It could contribute to existing assessment efforts on a campus, helping provide a deeper analysis, or it might be used to transform pedagogy in a more radical manner. It might also be used by students themselves, creating opportunities for holistic synthesis across both formal and informal learning activities”


• However……

“The potential for learning is clear, but the technology is still very young”
• Current technology is capable for applying ECD in GBL
• ECD is proven, but: few GBL applications and time-consuming
• More tools, R&D and ECD applications in GBL are needed
• Learning analytics (LA) seems promising to complement ECD
• LA might overload current technology for real-time interventions
• Future technology will provide new opportunities for embedded assessment blending ECD and LA for non-formal learning

Just do it…… because:

“serious games will only grow as an industry if the learning experience is definable, quantifiable and measurable” (Corti, 2011)

"If you are testing outside the game, you had better have a good reason for doing it. The very act of completing a game should serve as an assessment of whatever the intervention was designed to teach or measure". (Gee, 2009)
References – part 1


Collect & discuss Questions from the audience
(Live & via Twitter-backchannel #OEB12 #WSSGE)

Best practices & guidelines using games for learning

Assertion 1: Embedded Assessment is detrimental for gaming
Assertion 2: Embedded Assessment can(not) replace Certification because….
Assertion 3: For team based serious games - get the team to prepare individual presentations on their learning for the debriefing session (allow each team to work as a group).
Assertion 4: Performance measures are essential in assessing impact of games

Contact: rna@ou.nl
www.galanoe.eu
http://seriousgamessociety.org/
Hands-on session evaluating games

- (1) Shell hazard recognition (Igor, Arne)
- (2) Enercities (www.enercities.eu) (Rob, Jannicke)
- (3) Team up (Igor, Arne)
- (4) Metavals (Liz)
- (5) Siemens Plantville (www.plantville.com) (Johann)
- (6) First aid (http://first-aid-game.e-ucm.es) (Jannicke, Rob)
Hands-on session evaluating games

- (1) Shell hazard recognition (Igor, Arne) (validation phase)
  - single player, 3D, first person, action genre (TU Delft)
Hands-on session evaluating games

- (2) Enercities (www.enercities.eu) (Rob, Jannicke)
  - Build your own sustainable city
Hands-on session evaluating games

- (3) Team up (Igor, Arne) (used by Dutch police)
  - multiplayer (4 p), 3D, action genre (TU Delft/the Barn) for research and training of team collaboration and leadership)
Hands-on session evaluating games

- (4) Metavals (Liz)
  - collaborative game for learning statistical concepts

The MetaVals Game has 3 phases: Individual, correction and discussion and final answer. Let’s play the first phase!

**Individual**

You will have to classify six different items as Ratio, Nominal, Ordinal or Interval data.

**Correction**

You will see your partner’s answers to 6 different items and correct them by writing your own answers.

**Discussion**

You will discuss with your partner the 12 items and decide an accorded answer as a team.

*In all the phases, you will have a time-out and a Certainty Level Scale to write down your level of assurance when giving answer for each item.*
Hands-on session evaluating games

- (5) Siemens Plantville (www.plantville.com) (Johann)
  - Modernize three plants in Plantville
Hands-on session evaluating games

- (6) First aid (http://first-aid-game.e-ucm.es) (Jannicke, Rob)
  - Provide first aid