THE LEARNING PLAYER ASSESSMENT IN SERIOUS GAMES

Pradeepa THOMAS BENJAMIN
Phd Student

Lip6, Laboratore d'Informatique de ParisVI
UPMC, Université Pierre et Marie Curie
Paris, France
Pradeepa.thomas@lip6.fr

This project has been funded with support from the European Commission.

Sommaire

The Learning Player Assessment in Serious Games

I. Introduction

Simulations are often used in learning programs. Indeed, simulations allow the learner to test his decisions and adapt his strategy by analyzing the response of the system. Many companies use simulation-based serious games in order to train their employees.[1] At school, business games (based on an economic simulation) are used to teach the main principles of management or accounting for example. Through this method the players have roles of managers or presidents of companies. “But, this is not sufficient to be sure that the students will improve their knowledge and their understanding of the management of a company. Pedagogical functionalities such as explanations or diagnosis of what they have done must be added to the simulation software to create the conditions of learning” (J.-M. Labat, 2008). This is especially important that the main objective of serious game is learning (Prensky, 2003). The main task of the game design is to foster motivation and engagement of the learner. The underlying assumption is that the more the learner replays, the more he improves his knowledge. To verify this assertion, it is necessary to analyze the game traces of the player and assess his knowledge acquisition. For C.Conati, “Learning happens when the students actively build the connections between game and underlyling knowledge” (Conati & Manske, 2009).

In this paper we discuss the importance player tracking in serious games. Then we focus on use of tracking: the diagnosis of the player knowledge acquisition process. We conclude with our proposals of tracking indicators and feedback interfaces for both tutors and learners.

1 BNP Paribas with “StarBank the game” or Thalès with “The moonshield”.

II. Why should we track the player?

To follow the learner, we should trace his actions. We define the trace "as a temporal sequence of observed"(Settouti, Prié, Mille, & Marty, 2006). The trace analysis has been the subject of much research (Ngoc, Iksal, Choquet, & Klinger, 2009) (Connolly, Stansfield, Hainey, & Cousins, 2009)(Shute, Ventura, Bauer, & D Zapata-Rivera, s. d.). The data from the tracking in a serious game can be useful for different reasons. Firstly, the tracking is mainly used to evaluate the performance of the learner (Gee, 2003). But instead of evaluating the game, it is desirable to use these data to help progress in the game for example by giving clues to the player(Shute, Ventura, Bauer, & D Zapata-Rivera, 2009). The learner also needs to understand why he has lost, what his mistakes are and how to fix it. These indicators are also useful for the tutor.

Moreover, these elements can also help the designer to improve the game by highlighting errors in designs as is already the case for the learning systems (Marty, Heraud, Carron, & France,2004).
Finally, the tracking may also contribute to a dynamic adaptation of possible paths based on game actions of the learner (Conlan, Hampson, Peirce, & Kickmeier-Rust, 2009).

III. Learner assessment design

A. Domain knowledge representation

At first, in order to assess the knowledge gain of the player, we need to model the knowledge of the domain.

This raises the following questions: what is meant by knowledge? What knowledge do we want to model?

We define knowledge as a set of concepts that learners must master or control after or interaction with the learning system. Domain knowledge are those that must be mastered in full prior to the end of the game, some learners can master some aspects before playing. In this project, domain knowledge must be linked to the level 3 and 4 of the ECC. For instance, one item of this model can be: “understanding the role of the price on the demand function” linked to the level 5 of the ECC.

Nathalie DUCLOSSON in AMBRE-AWP (Nogry, Jean-Daubias, & Duclosson, 2004) presents a model of domain knowledge based on case-based reasoning. AMBER is an intelligent tutoring system which aims to help problem solving in mathematics. The model is enriched encountered with cases. In addition, Amber has a knowledge base (Chamade) which contains "two major groups of knowledge: those intended for communication with the learner, and those intended to diagnose learner responses and build help messages and explanation of errors.A problem database contains typical problems

already presented to the learner and the problems he has solved. "
In the context of serious games, we might well consider a similar architecture with a separate model of tutoring that exploits the results of a thorough analysis of traces of player interaction or even of all players.

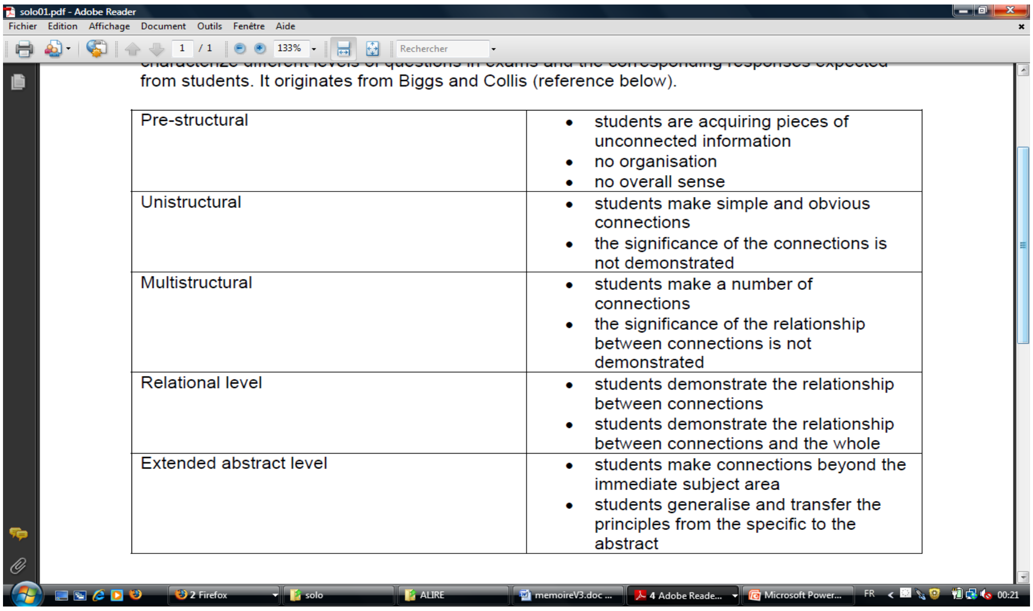
B. The learning objectives

Once the domain knowledge set, we need to define the educational objectives of the game and more specifically by the objectives of each level of the game
As such, we refer to the Taxonomy of Educational Objectives by Bloom (Bloom, 1956). It is a pedagogical model proposing a classification of levels of learning. More precisely that are the levels of thinking that Bloom and his colleagues consider as important in the learning process. It is assumed that the skills can be measured on a continuum from simple to complex. The taxonomy of educational objectives of Bloom is composed of six levels, including: knowledge, comprehension, application, analysis, synthesis and evaluation.

Figure Bloom taxonomy

Source :
http://www.fctl.ucf.edu/TeachingAndLearningResources/CourseDesign/BloomsTaxonomy/

Knowledge, comprehension, application and analysis can be assessed in a serious game. In fact, the more the game requires thought and involvement from the learner the more we advance in the taxonomy.



from students. It originates from Biggs and Collis (reference below).

Pre-structural	• students are acquiring pieces of unconnected information • no organisation • no overall sense
Unistructural	• students make simple and obvious connections • the significance of the connections is not demonstrated
Multistructural	• students make a number of connections • the significance of the relationship between connections is not demonstrated
Relational level	• students demonstrate the relationship between connections • students demonstrate the relationship between connections and the whole
Extended abstract level	• students make connections beyond the immediate subject area • students generalise and transfer the principles from the specific to the abstract

In the field of learning goals, we can also refer to the SOLO (Structure of the Observed Learning Outcome) (Biggs & Collis, 1982), taxonomy that describes the cognitive processes of learning on a scale of increasing complexity. It is used to characterize learning in increasing order of understanding. It can be used to set learning objectives and their evolution in the game.

Figure Levels of the SOLO taxonomy

This taxonomy could be used to classify the different goals of the game in each category and assign a different weight depending on the level of abstraction applied to the learner. This could allow us to define a model of the learner and possibly categorize the acquired knowledge and missing.

C. Learner assessment

The learner model represents what the player has learned but also his misconceptions. The learner model is a part of the domain model and also includes misconceptions (Barr, Beard, & Atkinson, 1976). When he plays, the learning player reveals his understanding of processes or mechanisms by the game actions he performs.

Our purpose in this paper is to assess the learner knowledge. The objective is to define the characteristics of the learner in terms of knowledge, skills and behavior. According to Stufflebeam, "assessment is the process which, describes, collects and provides useful information for judging decisions based on various possible"(Stufflebeam, 2002).

A skill is an "ability and capacity acquired through deliberate, systematic, and sustained effort to smoothly and adaptively carryout complex activities or job functions involving ideas (cognitive skills), things (technical skills), and/or people (interpersonal skills). See also competence"[2].

There are two main categories of evaluation: formative and summative. Formative assessment is to inform the tutor of the degree of mastery to correct and make necessary complements to the acquisition of skills. The main objective is to identify misconceptions and to correct them. This approach promotes metacognition showing the learner what he does not know help him to progress. The summative evaluation is used for verification and validation of skills. It focuses much more on the results. In serious game, the two types of evaluation can be used.

Stuflebeam proposes the CIPP (Contexte, Input, Process and Product) Model for Evaluation that can convey for business games. (Stufflebeam, 2002)

- Context : characteristics of the environment, goals and objectives, problems identification and diagnosis
- Input : resources idenfication, program designing
- Process : strategies and methods
- Product: evaluation of the outcomes and eventually loopback.

Many intelligent tutoring systems rely on "model tracing, based on the ACT theory (Anderson, 1983) and ACT-R (Anderson, Matessa, & Lebiere, 1997). This approach involves comparing step actions of the learner to those of the expert. The domain knowledge is represented by a system of rules like "if .. then .. . These rules associate the problem state actions and state changes that result. At each step in solving the problem, when a discrepancy is detected, the system triggers a feedback that requires the learner to go back on track. Simulations and virtual environments like Steve (Rickel & Johnson, 1998) also rely on rules defined by experts. To assist the learner Steve uses a predefined list of errors and associated behaviors to correct these errors. MASCARET (Baudoin, Beney, Chevaillier, & Le Pallec, 2007), allows to collect a lot of information about the behavior and actions performed. The analysis and interpretation of these data must be made by the tutor.

C.Despres and N.El-Kechai, as part of a monitoring system upon detection of failures at the SNCF, analyze the activity of the learner to identify discrepancies between what he did and reference model. (El-Kechaï & Després, 2006). The activity of the learner is projected onto a plan and then compared with the procedure model. They uses the classification of erroneous

2"http://www.businessdictionary.com/definition/skill.html actions of Hollnagel (Hollnagel, 1998). This classification distinguishes between observable performed actions (phenotypes) and the factors and processes which generated the action (genotypes). Part of this analysis could be applied to serious games. Indeed, some groups of simple phenotypes such as time, duration, a bad object, the sequence is also present in games. It would be interesting to classify the observed actions in the game according to the categories of Hollnagel. The authors offer the trainer, automatically, the causal mechanisms (genotypes) which may explain the occurrence of the phenotype. For this, it uses CREAM, a method of reliability analysis that allows human after several iterations to infer the causes of erroneous action. Each iteration provides an "effect" and a "cause" which in turn becomes the "effect" of the next iteration. The main difficulty in the context of a game is to be able to identify all possible causes of a wrong action to feed the model.

IV. Feedback

A. Examples

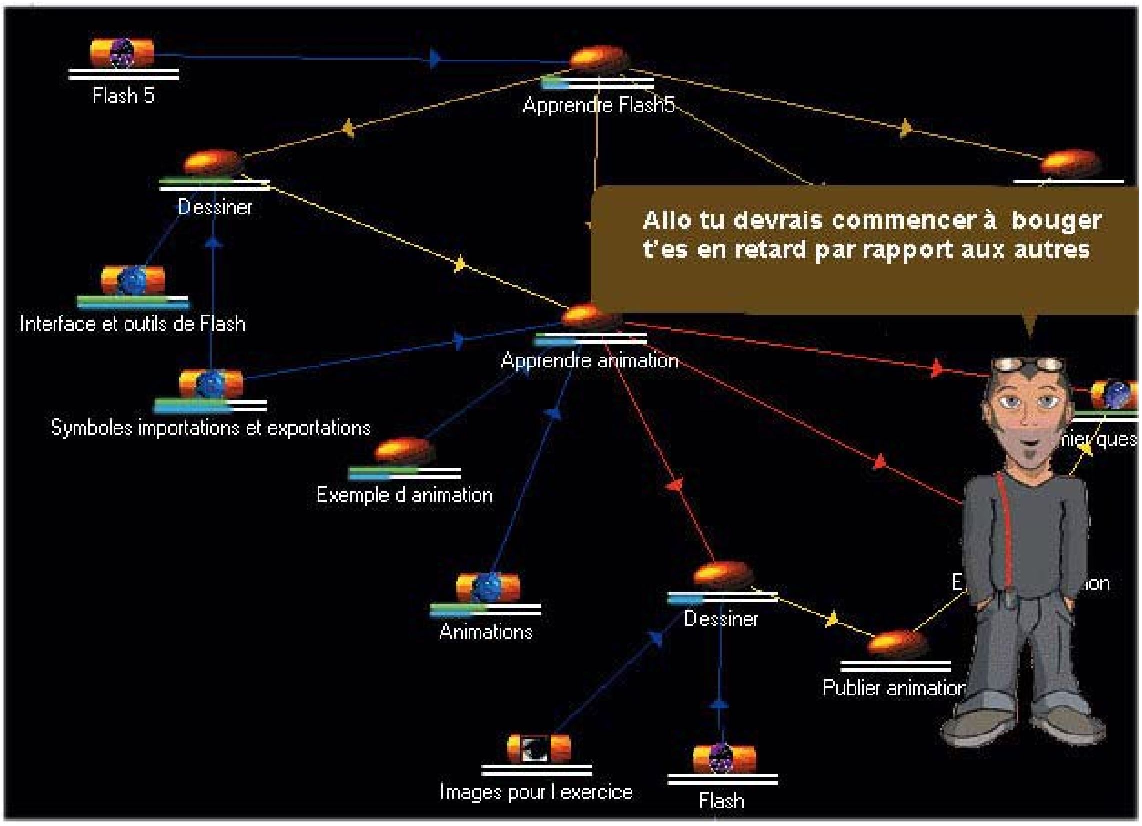
The feedback is one of the most important aspect of a serious game. In order to help the player to learn, the game should present to him accurate information(Azevedo & Bernard, 1995) (Person, Magliano, & A. C. Graesser, s. d.).

We should consider different aspects of the feedback in serious games by answering the following questions: when? What information? Which format?
If we consider a formative evaluation process, feedback can be immediate and possibly punctuated with explanations to help the learner understand the causes of his error. Feedback can also be carried out at the end of a level or the whole game to allow enough freedom to the player and not interrupting immersion in the game. The benefit of a return at the end of the game is that the learner can correct his mistakes himself and therefore return to a proper resolution of the problem which is posed.

D. McNamara in (Jackson, A. Graesser, & McNAMARA, 2009) distinguishes information-based, consequence-based, and point-based feedback. “ Information-based feedback takes the form of providing specific feedback on the accuracy and quality of answers and actions within the ITS or game. Consequence-based feedback occurs when the system reacts to the user's responses or actions by changing the system path. For example, when a student answers incorrectly in an ITS, the student may be redirected into remediation. When a player is killed in a game, the game ends. This is the ultimate form of feedback. Both information-based and consequence-based feedback are common to most ITSs. By contrast, point-based feedback is a feature more characteristic of games and is less common in ITSs. Point-based feedback can be conveyed in the form of cumulative points, progress bars, and levels. All of these are based directly or indirectly on the accuracy or the quality of responses and actions by the user but they differ in their specificity and detail. A progress bar is a means of conveying performance schematically (e.g., shaded progress bar, skillometers*),* without the specificity of points. Likewise, the use of levels is less specific than points and conveys feedback on performance across a longer time-span.”

In ExploraGraph, an intelligent tutor to facilitate navigation in a hypermedia, A. Dufresne and S. Prom Tep, define a learner model represented as a directed graph with a progress indicator for each stage of learning. (Dufresne, 2006)

Figure The feedback of exploragraph



D. Zapata-Rivera in (Diego Zapata-Rivera, 2010) considers using assessment to guide the learner: in ABLE (Assessment-Based Learning Environment), a character named Dr. Grammar provides adaptive instructional feedback (i.e., rules, procedures, examples and definitions) based on the student model.

C.Conati and X.Zhao in Prime Climb, a game to teach number factorization, their goal is to accurate hints for each student (Conati & Zhao, 2004). In the game, the player is assessed by a probabilistic student model. An agent helps the player during the game by relying on a probabilistic model of his knowledge at a defined moment. The student model evolves with the progression of the player in the game. The player can ask the agent for help and the agent gives hints. But one the specificity of this game is that the agent can automatically decide to intervene in the game. Indeed, the agent gives hints when the student model (based on probabilistic values) indicates that major pieces of knowledge have been missed. If the agent doesn't intervene, we can consider that the player is not really able to move on in the game because he lacks some prerequisites.

B. Our proposal

1. <u>For the learner</u>

In this business game we propose feedbacks at the end of the game, when the player has lost. We refer here to our work on StarBank the game (Thomas, 2010).

The feedbacks consist of three screens:

- a screen showing the results of the player with the accomplishment or not of the learning objectives of the level and the distribution of the activities with their profitability.
- a display board analyzing the causes of failure

To better reflect the information provided to the player, we decided use current color codes: a validated objective or a profitable activity are marked in green, failures or unprofitable activities in red.

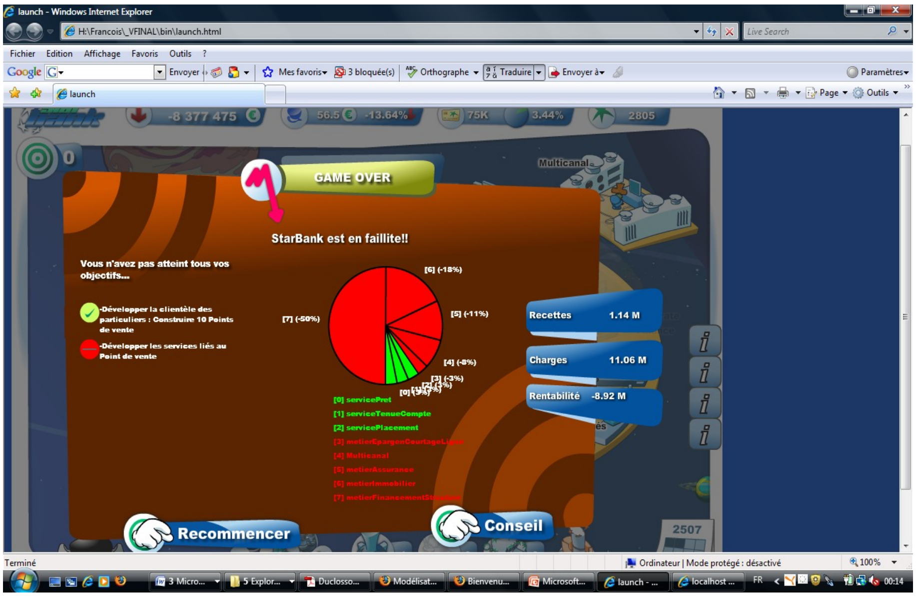
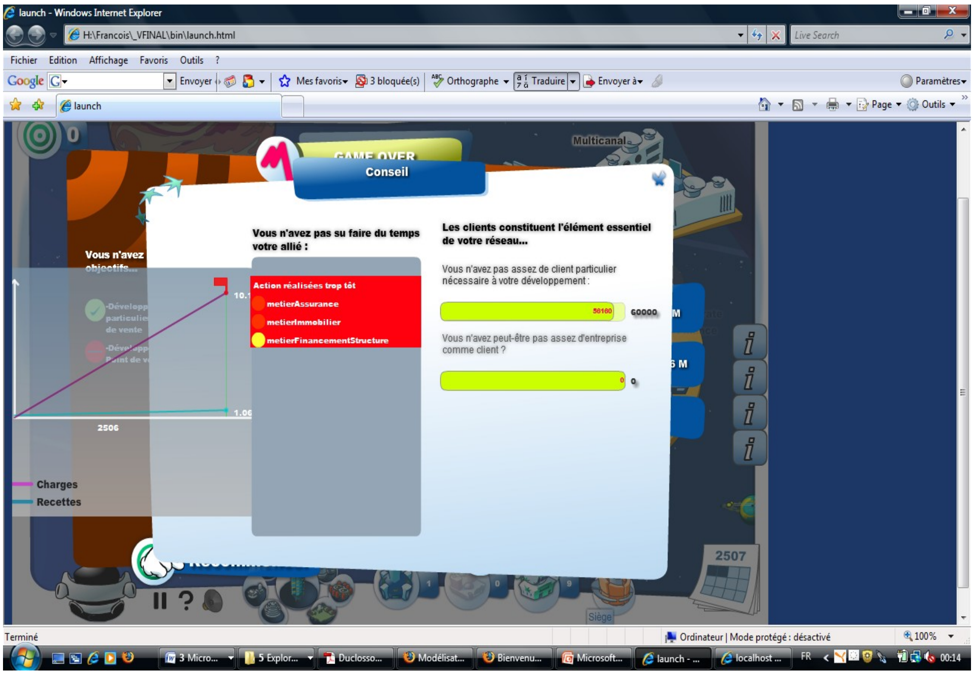


Figure Screen shots of StarBank the game



Regarding the precocity of actions, we turn from yellow (a little early) to orange (early) to red (very early). By clicking on each activity prematurely, the player has the opportunity to show the effect on the profitability of his decision.
On the right, he can also compare his performance in terms of clients over the required threshold.

We also suggest the following indicators based on our work in (Thomas, Yessad, & J. M. Labat, 2011) :

- erroneous actions, are those that do not contribute to the progression in the game,
- forbidden actions are actions that do not comply with the game rules,
- premature actions are those happening too early,
- belated actions are those happening too lately,
- inaccurate actions that can be considered as erroneous actions,
- unnecessary actions are actions that do not impact negatively on the progress of the game but do not contribute to accomplishing the objectives,
- imprecise actions contribute to progress in the game without fully meet the learning objectives; the multiplication of these actions by the learner reveals a poor mastering of the domain and approximate knowledge, the player manages to sidestep the problems without using the relevant solution.

We can also add for a business game the following indicators:

- costly activities
- lacking investment activities
- non efficient choices (ex : the combination of production factors)
- satisfaction of the stakeholders (clients, employees, managers, banks, suppliers…)
- behavior and reaction to external shocks
- accuracy of the advertizing policy
- accuracy of the price policy

2. For the tutor

The tutor needs different types of feedbacks: about each student, about all the students and the progression between many game sessions.

First, can get the same indicators as each student.

Then we suggest the following indicators:

- the average time per level of play
- errors most frequently committed
- the sticking points and associated knowledge
- points of great success and the knowledge associated
- the best performance
- the worst performance
- details of the strategies on all points of view (commercial, production, marketing)
- strengths and weaknesses of each student

V. Conclusion

Our contribution in this paper was to provide an overview on different evaluation methods in serious games. Indeed, it is possible to assess the player in real time synchronously by giving immediate feedbacks. But with the risk of damaging the immersion mechanism created by the game. It is also possible to generate feedbacks at the end of the game. We should not only highlight the missing elements and misconceptions of the learner but also to enhance it by showing what he can do well.

The tutor needs in turn, aggregated indicators to situate the overall level of their class. Evaluation can also be used for real-time adaptation in the game in order to offer the player a context suited to their abilities and previous mistakes.

VI. References

Anderson, J. R. (1983). *The architecture of cognition* (Harvard University Press.).

Anderson, J. R., Matessa, M., & Lebiere, C. (1997). ACT-R: a theory of higher level cognition and its relation to visual attention. *Human-Computer Interaction*, *12*, 439–462. doi:http://dx.doi.org/10.1207/s15327051hci1204_5

Azevedo, R., & Bernard, R. M. (1995). A Meta-analysis of the Effects of Feedback in Computer-Based Instruction. *Journal of Educational Computing Research*, *13*(2), 111-27.

Barr, A., Beard, M., & Atkinson, R. C. (1976). The computer as a tutorial laboratory: the Stanford BIP project. *International Journal of Man-Machine Studies*, *8*(5), 567-582, IN1-IN2, 583-596. doi:10.1016/S0020-7373(76)80021-1

Baudoin, C., Beney, M., Chevaillier, P., & Le Pallec, A. (2007). Recueil de traces pour le suivi de l'activité d'apprenants en travaux pratiques dans un environnement de réalité virtuelle. *STICEF*, *14*, 389-418.

Biggs, J. B., & Collis, K. F. (1982). Evaluating the quality of learning : the SOLO taxonomy (Structure of the observed learning outcome) / John B. Biggs, Kevin F. Collis. New York : Academic Press. Consulté de http://platon.serbi.ula.ve/librum/librum_ula/ver.php?ndoc=232692

Bloom, B. S. (1956). *Taxonomy of Educational Objectives, Handbook 1: Cognitive Domain* (2e éd.). Addison Wesley Publishing Company.

Conati, C., & Manske, M. (2009). Evaluating Adaptive Feedback in an Educational Computer Game. Dans Z. Ruttkay, M. Kipp, A. Nijholt, & H. H. Vilhjálmsson (Éd.), *Intelligent Virtual Agents* (Vol. 5773, p. 146-158). Berlin, Heidelberg: Springer Berlin Heidelberg. Consulté de https://acces-distant.upmc.fr/http/www.springerlink.com/content/y1p6l46183727716/

Conati, C., & Zhao, X. (2004). Building and evaluating an intelligent pedagogical agent to improve the effectiveness of an educational game. *Proceedings of the 9th international conference on Intelligent user interfaces*, IUI '04 (p. 6–13). New York, NY, USA: ACM. doi:10.1145/964442.964446

Conlan, O., Hampson, C., Peirce, N., & Kickmeier-Rust, M. (2009). Realtime Knowledge Space Skill Assessment for Personalized Digital Educational Games (p. 538-542). Présenté à Advanced Learning Technologies, 2009. ICALT 2009. Ninth IEEE International Conference on.

Connolly, T., Stansfield, M., Hainey, T., & Cousins, I. (2009). Arguing for Multilingual Motivation in Web 2.0: a Games-Based Learning Platform for Language Learning. Présenté à European Conference on Games Based Learning, Graz, Austria.

Dufresne, A. (2006). Explor@Graph the Representation Power of Adaptive Conceptual Graphs. *Advanced Learning Technologies, IEEE International Conference on* (Vol. 0, p. 1111-1112). Los Alamitos, CA, USA: IEEE Computer Society. doi:http://doi.ieeecomputersociety.org/10.1109/ICALT.2006.168

El-Kechaï, N., & Després, C. (2006). A plan recognition process, based on a task model, for detecting learner's erroneous actions. *Lecture Notes in Computer Science* (Vol. 4053, p. 329-338). Présenté à Intelligent tutoring systems, eight international conference, Jhongli, Taiwan: Springer Berlin.

Gee, J. P. (2003). What video games have to teach us about learning and literacy. *Comput. Entertain.*, *1*(1), 20-20.

Hollnagel, E. (1998). Cognitive Reliability and Error Analysis Method (Elsevier Science.). Oxford.

Jackson, G. T., Graesser, A., & McNAMARA, D. (2009). Intelligent tutoring and games (ITaG). *AIED 2009 14 th*, 44-65. doi:10.4018/978-1-61520-713-8.ch003

Labat, J.-M. (2008). Designing Virtual Players for Game Simulations in a Pedagogical Environment: A Case Study. Dans Z. Pan, X. Zhang, A. Rhalibi, W. Woo, & Y. Li (Éd.), *Technologies for E-Learning and Digital Entertainment* (Vol. 5093, p. 487-496). Berlin, Heidelberg: Springer Berlin Heidelberg. Consulté de https://acces-distant.upmc.fr/http/www.springerlink.com/content/a433547tv7205466/

Marty, J. C., Heraud, J.-M., Carron, T., & France, L. (2004). A quality approach for collaborative learningscenarios. *Learning Technology Newsletter of IEEE*, *6*(24).

Ngoc, D. P. T., Iksal, S., Choquet, C., & Klinger, E. (2009). UTL-CL: A Declarative Calculation Language Proposal for a Learning Tracks Analysis Process. *Proceedings of the 2009 Ninth IEEE International Conference on Advanced Learning Technologies* (p. 681-685). IEEE Computer Society. Consulté de http://portal.acm.org/citation.cfm?id=1582708.1582881

Nogry, S., Jean-Daubias, S., & Duclosson, N. (2004). ITS Evaluation in Classroom: The Case of Ambre-AWP. Dans J. C. Lester, R. M. Vicari, & F. Paraguaçu (Éd.), *Intelligent Tutoring Systems* (Vol. 3220, p. 511-520). Berlin, Heidelberg: Springer Berlin Heidelberg. Consulté de http://www.springerlink.com/content/k8regh5v62fbr8dw/

Person, N. K., Magliano, J. P., & Graesser, A. C. (s. d.). Collaborative Dialogue Patterns in Naturalistic One-to-One Tutoring. *Applied Cognitive Psychology*, *9*(6), 495-522. doi:10.1002/acp.2350090604

Prensky, M. (2003, octobre). Digital game-based learning, *1*(1).

Rickel, J., & Johnson, W. L. (1998). STEVE: A Pedagogical Agent for Virtual Reality. *IN PROCEEDINGS OF THE SECOND INTERNATIONAL CONFERENCE ON AUTONOMOUS AGENTS*, 332--333.

Settouti, L., Prié, Y., Mille, A., & Marty, J. C. (2006). Systèmes à base de traces pour l'apprentissage humain. Présenté à TICE.

Shute, V. J., Ventura, M., Bauer, M. I., & Zapata-Rivera, D. (2009). Melding the Power of Serious Games and Embedded Assessment to Monitor and Foster Learning: Flow and Grow. *The Social Science of Serious Games: Theories and Applications* (PA: Routledge/LEA., p. 293-319). Philadelphia: U. Ritterfeld, M. J. Cody, & P. Vorderer. Consulté de http://scholar.google.com/scholar?hl=en&q=%E2%80%A2+Valerie+J.+Shute%2C+Matthew+Ventura%2C+Malcolm+Bauer%2C+and+Diego+Zapata-Rivera&btnG=Search&as_sdt=2000&as_ylo=&as_vis=0

Shute, V. J., Ventura, M., Bauer, M. I., & Zapata-Rivera, D. (s. d.). Melding the Power of Serious Games and Embedded Assessment to Monitor and Foster Learning: Flow and Grow. *The Social Science of Serious Games: Theories and Applications* (PA: Routledge/LEA.). Philadelphia: U. Ritterfeld, M. J. Cody, & P. Vorderer. Consulté de http://scholar.google.com/scholar?hl=en&q=%E2%80%A2+Valerie+J.+Shute%2C+Matthew+Ventura%2C+Malcolm+Bauer%2C+and+Diego+Zapata-Rivera&btnG=Search&as_sdt=2000&as_ylo=&as_vis=0

Stufflebeam, D. L. (2002). The CIPP Model for Evaluation. Dans D. L. Stufflebeam, G. F. Madaus, & T. Kellaghan (Éd.), *Evaluation Models* (Vol. 49, p. 279-317). Dordrecht: Kluwer Academic Publishers. Consulté de https://acces-distant.upmc.fr/http/www.springerlink.com/content/pl14821x87334086/

Thomas, P. (2010). Le suivi de l'apprenant dans un serious game au sein d'une plateforme d'apprentissage. Présenté à Rencontres Jeunes Chercheurs EIAH, Lyon.

Thomas, P., Yessad, A., & Labat, J. M. (2011). Petri nets and ontologies : tools for the « learningplayer » assessment in serious games. *Proceedings of the 11th IEEE International Conference on Advanced Learning Technologies*. Présenté à International Conference on Advanced Learning Technologies, Athens, Georgia, USA: IEEE Computer Society.

Zapata-Rivera, Diego. (2010). Adaptive, Assessment-Based Educational Games. Dans V. Aleven, J. Kay, & J. Mostow (Éd.), *Intelligent Tutoring Systems* (Vol. 6095, p. 435-437). Berlin, Heidelberg: Springer Berlin Heidelberg. Consulté de https://acces-distant.upmc.fr/http/www.springerlink.com/content/e1731r682k3h761x/