The Challenge of Assessing Learning in Open Games: HORTUS as a Case Study

Dr. Franziska Spring, Dr. James W. Pellegrino,
Learning Sciences Research Institute, University of Illinois at Chicago, United States,
Email: franspring@gmail.com , pellegjw@uic.edu

Abstract

Educational designers and researchers are faced with new challenges when it comes to the design of open games for learning that allow players to choose their own solution path through the game. The goal of this empirical study with 70 university students was to be able to tell at any point in the game what the player knows and whether a specific learning goal has been attained. The analysis of learning success considered two performance categories: continuous improvement and learning by failure. An in-game assessment and an external posttest served as verification of a player's hypothesized state of learning. Behavioral patterns extracted from game playing provided evidence of a player's learning success or failure and were shown to have predictive accuracy.

Introduction

Researchers agree that games often provide deep conceptual and meaningful learning (Gee, 2003; Shaffer, 2006; Squire, 2008). However, little is known of how original games provide effective learning and how they effect learners (Squire, 2008). This paper concentrates on so-called open games for learning and begins by considering what they are and how players learn in such games. This is followed by a discussion of the state of the art regarding assessment of learning in games. The majority of the paper presents a case study of learning in the game *Hortus*. The game's learning goals are described, as well as how they are assessed, followed by a discussion of empirical results. We conclude with an outlook regarding future prospects and challenges in this area.

Open Games and Learning

Open games are similar to simulations, as they provide an open environment where learners are loosely guided and can choose among multiple pathways. The most popular examples of this genre applied to an educational context are commercial games like *Civilization* or *Sim City*. There are a variety of goals that have to be attained in order to win the game. Learning is regarded as the understanding of systems and their dynamics. Learners are encouraged to experiment and to experience causalities of a virtual world. The strengths of open games are that they feature layers of complexity that promote curiosity, discovery, and replayability—important characteristics for intrinsic motivation (Malone, 1987).

There are only a few studies that investigate how people learn in open games and simulations. Usually, the nature of the game dictates learning goals to a certain degree and open games cover a variety of such goals. They include conceptual understanding, science process skills, or scientific discourse (Honey and Hilton, 2011). The most popular learning goal is conceptual learning (Honey and Hilton, 2011; Gee, 2005; Squire, 2008). Conceptual learning involves gaining an understanding and applying rules and concepts of a system, such as the functioning of machines, the human body, or horticultural systems. Learning within games involves many processes such as learning through failure, learning by practice, and learning through competition (Gee, 2003, 2005; Squire, 2008, 2011). Learning by failure and the practice principle are our focus because they are applicable to learning in *Hortus*.

Failure takes on a different meaning in a playful learning environment. If players fail in a game, the worst that could happen is the death of their character or the destruction of their city. But after resurrecting the character or rebuilding the city, they can try again. Learning in games tolerates failure without real punishment. This allows trying out different solution paths and going to the limit of the system while remaining on the safe side without any real-world consequences.

Practice is very important in games. Games provide situations and opportunities where players have the chance to practice their recently learned skills. Other than in drill and practice games, these situations are embedded into the story of the game. Suddenly, practicing skills until they become routine is not boring anymore, but rather meaningful in a situated context. Players can achieve a mastery level for certain skills and new skills have to be learned or combined with old ones to meet the next challenge.

Assessment of Game-Based Learning

Assessment of learning from games is usually conducted apart from the game using traditional methods such as multiple choice questions (Burgos, 2008; Tan & Biswas, 2007; Wang, 2008) or qualitative interviews (Galarneau, 2005; Squire, 2008; Squire & Durga, in press). Shaffer (2009) assesses epistemic games with so-called epistemic frame inventories (EFI). In epistemic games, players take over the role of professionals in the respective field. This rather new assessment concept deals with a variety of methods to assess players in a specific knowledge domain. For instance, players have to create concept maps or they are asked to use their learned skills to solve designed tasks. Those methods, however, are not integrated in the game environment. This can be contrasted with the assessment approach in the project “SimScientists” (Quellmalz et al., 2009a, 2009b) which has the goal of encouraging model-based reasoning in science. Students learn about food chains in simulation environments with embedded interactive tasks. In the context of the simulation, students have to categorize fish in the hierarchy of the food chain and their decisions are immediately assessed.

There are few cases of knowledge assessments seamlessly integrated into a game. Furthermore, such assessments are usually highly structured so that learners have to take the same paths as other learners. In open games, learners are not restricted by given paths. Learning progress in open games is defined as a change in a learner’s reaction to a certain event or situation in the game and it is difficult to monitor or to control. This reaction is manifest as either an increase or decrease in performance. The challenge is to develop approaches to assessment that respect the characteristics of open games and do not interrupt play flow. We next describe a case study of attempting to do so in the open game *Hortus.*

Figure 1. Hortus – Level 2.

Hortus – A Case Study

Hortus is an online strategy game specifically designed for research on open games and assessments. It is turn-based which means that most of the progress in the game is only visible after the player actively clicks a 'Next Day' button. The player takes over the role of an herbalist who has to create a potion to heal the sick people in the village. Therefore, players have to plant herbs in their garden plot, grow them, and harvest leaves for the potion. The game consists of four levels. The first level is a tutorial where players learn the basic skills of herb growth such as planting, watering, and harvesting. In levels 2 and 3, new herbs are introduced (see Figure 1). Those herbs interact with each other either positively (symbiotic) or negatively (competitive). Level 4 provides a starting situation and serves as an in-game assessment. Each player has the same goals and the same number of days in which to finish a given level and it is up to the player how quickly and efficiently they reach the game goal.

Learning Goals

The player has to learn six different plant interactions that are produced when placing different kinds of herbs next to each other (see Figure 2). The visual display indicates how well the player is using those relations and feedback occurs in various ways. An overall score rates how well a player takes advantage of those interactions. Within each turn, the rating changes and indicates an improvement or decrease in the performance of the player. The score is only kept for one level and restarts when moving to the next level. When herbs are placed next to each other on the plot, the player receives immediate visual feedback in the form of tiles that represent the nature of the interaction. Players have to build up their own hypotheses from this information as to the quality of the interaction. Another source of information is the number of leaves growing on each herb. A positive plant interaction accelerates leaf growth drastically while a negative interaction stops or even removes leaves from the herbs.

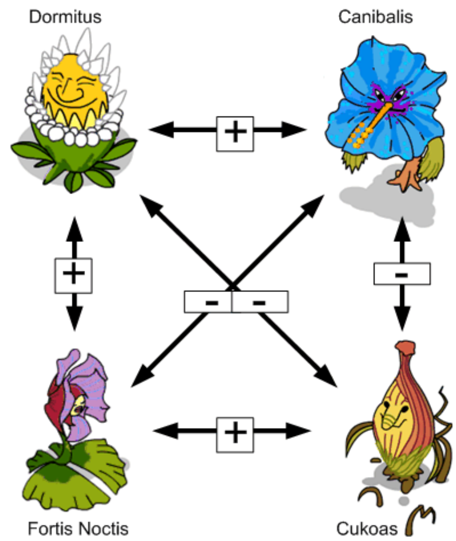

***Figure 2*. Four plants with six different interactions**

Verification of Learning Goals

A player's learning, i.e., was a specific plant interaction understood?, can only be assessed in an indirect way by analyzing performance patterns. It is assumed that players with good strategies, meaning a short number of turns and a high number of positive interactions, have understood most of the plant interactions because they are able to apply their knowledge in the form of a strategy. For a more detailed assessment of specific learning goals, they are divided into two categories: improving performance (positive interactions) and learning by failure (negative interactions). It is assumed that a player has to experience all plant interactions in order to learn them. This is also true for negative interactions. The more positive interactions a player has, the more likely it is that he understood each interaction and is able to use that knowledge in developing a play strategy. Improving a player's performance is a modified version of Gee's practice principle. For negative interactions, the player has to experience each at least once but then has to avoid them. There were two kinds of assessments designed to verify the assumptions above. First, there is an in-game test scenario (Level 4). Figure 3 shows the starting situation the player has to modify. It is a particularly problematic situation and the assessment is whether all the negative interactions (dotted tiles) disappear and the positive tiles (wavy tiles) increase in number in the first turn. As a second verification, there is an external posttest in which players have to solve different tasks that are related to the learning goals in the game.

***Figure 3*. Level 4 – Test scenario with a bad starting situation**

The Participants

153 college students were recruited for the online study: 22 from the United States, 117 from Austria, and the remainder from other countries. Their ages ranged from 18 years to 61 years. For analysis purposes, a valid data set required that each player finish Level 4 in his or her first attempt. 70 participants met this criterion and their data contributed to the analysis corpus. The online experiment included playing the game—approximately 45 minutes—and taking the posttest—approximately 5 minutes.

Results and Discussion

The player's overall game score, summarized across all negative and positive interactions, increased as the player used more positive interactions. However, the score did not reveal how many times a player used particular plant relations—only how many were used in total. Thus, the information derived from the overall game score could only indicate an increase or decrease in the use of positive and negative relations in general, and is only a gross indication of the progress of learning. Therefore, to more precisely assess learning, the categories of positive and negative interactions were analyzed separately to determine if a specific interaction was "known" and being used in game play.

Analysis of Positive Interactions

Players need to learn about plant interactions that are symbiotic and cause plant leaves to grow faster. The leaves are necessary for the potions to heal the sick villagers in the game, which is the desired end state. The more symbiotic relations a player has in their garden plot the faster the game goal is achieved. As mentioned before, players have to experience specific positive interactions in order to understand them. However, the challenge was to find out how many experiences of this kind were necessary so that a “hypothesis” could be generated that a relation was learned. Eventually, for every player, a so-called ‘breakthrough moment’ was identified. This moment is detected by searching for the current maximum of positive interactions on the field. The maximum has to be equal to or higher than the median value over all players’ maximum of positive interactions for the current level.

For example, the plant relation Dormitus – Canibalis had to be learned in level 2. The median value for this relation was 4 for a learning breakthrough. Table 1 shows a player (ID 62) with a breakthrough moment at Turn 6, with 8 positive interactions—well above the median value. The same player showed a similar strong pattern throughout for all plant interactions. However, not every player had such a clear behavior pattern. Some players put a strong emphasis on only specific plant interactions while neglecting others. It is assumed that a player’s strategy plays an important role in this behavior.

Table1: Player ID 62 shows a learning breakthrough from turn 5 to 6.

Can_Dorm	Level	2	2	2	2	2	2	2	2	2	2	2	2	2	3	3	3	3	3	4	4	4
	NoOfTimesReplayLevel	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Turn	0	1	2	3	4	5	6	7	8	9	10	11	12	0	1	2	3	4	0	1	2
	NoOfPosInteractPerTurn	0	0	0	0	0	0	8	8	8	8	8	8	8	5	4	4	5	5	5	5	5

In Level 4, players had to re-organize a given situation that contained many negative plant relations (see also Figure 3). In our example, the criterion to pass level 4 for the relation Dormitus – Canibalis required that there be at least 3 positive interactions in the first turn (median value for this plant relation). Thus, player ID 62 passed this test with 5 such interactions in the first turn. The small number of turns in level 4 (3 turns out of 10 possible) also shows use of a good strategy overall.

When analyzing all the positive interactions over all players, there was a 70% match between the game results (met/not met the criterion) and the Level 4 test scenario results (passed/failed). This means that learning success or failure at Level 4 for a given learning goal can be predicted with a 70% probability based on play information from level 2 or 3. The predominant tendency was over-prediction of learning. While this is an adequate level of prediction, one would ultimately like a higher degree of match.

Negative Interactions

The negative plant interactions were learned through experiencing them and then showing avoidance. A player was assumed to have understood a negative interaction when there was a breakthrough moment followed by no further negative interactions—called the ‘flat phase’. The breakthrough moment in this context is defined as a temporary increase in the number of negative interactions. In the best case, a player only has one breakthrough moment followed by a flat phase until the end of the game. However, this kind of behavior has to be assessed in each level of the game. Someone with several breakthroughs and hardly any flat phases in between is not likely to have achieved the learning goal. Therefore, the criterion for passing the Level 4 test scenario is quite strict. The player was not allowed to have any negative interactions in the first two turns. It is expected that if a player acquired this knowledge, he would remove all the existing negative interactions and also try to avoid them in the future. The example in Table 2 shows a player who had a successful negative breakthrough followed by a flat phase until Level 4.

Table 2: Player ID 62 reduced the five negative interactions to zero.

Dorm_Cuk	Level	2	2	2	2	2	2	2	2	2	2	2	2	2	3	3	3	3	3	4	4	4
	NoOfTimesReplayLevel	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Turn	0	1	2	3	4	5	6	7	8	9	10	11	12	0	1	2	3	4	0	1	2
	NoOfNegInteractPerTurn	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 3 shows a player who did not meet the learning goal. He had several peaks with a high number of negative interactions but no real flat phase. From the eight given negative interactions in the starting scenario of Level 4, he left half of them on the field. When analyzing all the negative interactions, the data revealed a 83% match between the game results (met/not met) and the test scenario results (passed/failed). This means that a learning success or failure can be predicted with 83% probability based on performance in level 2 or 3. Clearly, prediction in this case was superior to positive interactions.

Table 3: Player ID 156 had continuously negative relations.**Error! Not a valid link.**

Dorm_Cuk	Level	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	4	4	4
	NoOfTimesReplayLevel	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	0	0	0
	Turn	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	0	1	2	0	1	2	3	4	5	6	7	0	1	2
	NoOfNegInteractPerTurn	1	1	1	1	2	2	1	3	3	4	3	3	3	2	2	4	4	2	2	2	2	0	2	1	1	1	4	4	1

Posttest

Initially, a knowledge posttest was developed, designed to verify learning results from the game. Although the test questions seemed to be a fair assessment of the learning goals, the results were not promising when compared with the game data. Only 50% of the cases matched the game results. Players who used many positive plant interactions still gave incorrect answers in the posttest. The opposite also occurred—weak game players answered the posttest questions correctly. An explanation for this incompatibility may be that questions in the posttest require a different kind of knowledge than that learned in the game. In the game, players receive immediate feedback when they place plants on the field. In the test, players had to actively associate interaction type with the plants. While players might be aware of the different interactions and plant associations as part of sophisticated game playing strategies, they might not be able to apply or transfer this knowledge to a more abstract environment as represented by problems presented in the posttest. Thus, their knowledge may be highly situated in the game that circumscribes transfer. This clearly is a matter that needs further investigation.

Conclusions and Future Directions

The game score reflected how well a player took advantage of the plant interactions, but not if each relation was understood. By categorizing the negative and positive interactions into classes of learning goals such as ‘learning by failure’ and ‘learning by improving performance’, respectively, it was possible to abstract behavioral patterns that indicated successful learning of specific information. In addition, it was easier to detect if a player learned the negative interactions by avoidance than analyzing a continuously improving strategy based on positive interactions. When analyzing all the positive interactions, the data revealed a 70% match between the game results (met/not met) and the test scenario results (passed/failed). In contrast, the negative plant interactions revealed an 83% match.

Educational designers and researchers are confronted with many issues when it comes to designing open games for learning. The chasm between ‘learning’ and ‘fun and motivating’ is difficult to bridge. First, there is the challenge to walking a fine line between developing an exciting game and designing learning goals that do not compete with the game goals. For instance, a player’s strategy in *Hortus* might result in a weak and unclear behavioral pattern where it is not possible to see if the learning goals were attained. Second, there is the dilemma of setting restrictions on choices in the game to enable the assessment of learning. A game with confined choices makes it easier to assess learning than one where players are free to explore but it also runs counter to the nature of open games. These two issues require further investigation and analysis.

Simulations and complex game-like learning environments are growing in use and importance in our society (e.g., Honey & Hilton, 2011). With the improvement of technology and increasing experience with these kinds of learning environments, more complex learning scenarios should be possible that can enhance the nature and quality of learning. However, assessment methods have to be aligned with the learning environment to support the effectiveness of such learning environments.

References

Burgos, D., Moreno-Ger, P., Sierra, J., Fernandez-Manjon, B., Specht, M., & Koper, R. (2008). Building adaptive game-based learning resources: The integration of ims learning design and <e- adventure>. *Simulation Gaming*. 39(3), 414-431.

Galarneau, L. (2005). Authentic learning experiences through play: Games, simulations and the construction of knowledge. Paper presented at the DiGRA, Vancouver, Canada.

Gee, J. (2003). *What video games have to teach us about learning and literacy*. New York: Palgrave Macmillan.

Honey, M. A. & Hilton, M. H. (2011) Learning Science: Computer Games, Simulations, and Education. National Research Council. Washington, DC: The National Academies Press.

Malone, T. W. & Lepper, M. R. (1987) Making learning fun: A taxonomy of intrinsic motivations for learning. In R. E. Snow & M. J. Farr (Eds.). *Aptitude, learning, and instruction. III: Conative and affective process analysis*. Hillsdale, NJ: Lawrence Erlbaum Associates, pp. 223-253.

Quellmalz, E. S., Timms, M., & Buckley, B. (2009a) Using science simulations to support powerful formative assessments of complex science learning. *WestEd*.

Quellmalz, E. S., Davenport, J., Timms, M., & Buckley, B. (2009b) Quality science simulations for formative and summative assessment. *WestEd*.

Shaffer, D. W., Hatfield, D., Svarovsky, G. N., Nash, P., Nulty, A., Bagley, E., et al. (2009) Epistemic network analysis: A prototype for 21st century assessment of learning. *International Journal of Learning and Media*, 1(2), 33-53.

Shaffer, D. (2006). *How computer games help children learn*. New York: Palgrave Macmillan.

Squire, K. (2011). *Video games and learning: Teaching and participatory culture in the digital age*. New York: Teachers College Press.

Squire, K., & Durga, S. (in press). Productive gaming: The case for historiographic game play. To appear in R. Ferdig (Ed.) *The handbook of educational gaming*. Hershey, PA: Information Science Reference.

Squire, K. (2008). Open-ended video games: A model for developing learning for the interactive age. In K. Salen (Ed.) *The ecology of games: Connecting youth, games, and learning*. (167-198) Cambridge, MA: The MIT Press.

Tan, J. & Biswas, G. (2007) Simulation-based game learning environments: Building and sustaining a fish tank. In: *Proceedings of the First IEEE International Workshop on Digital Game and Intelligent Toy Enhanced Learning*, Jhongli, Taiwan, pp. 73–70.

Wang, T. H. (2008) Web-based quiz-game-like formative assessment: Development and evaluation. *Computers and Education*. 51, 1247–1263.