Theory Informed Learning Analytics for Embedded Assessment in Game-based Science Learning: Towards Actionable Intelligence

Wanli Xing

[Serious games] will not grow as an industry unless the learning experience is definable, quantifiable and measurable. Assessment is the future of serious games. (Ritterfeld, Cody & Vorderer, 2009)

Science not as a noun . . . but as a verb, a process, a set of activities, a way of proceeding and thinking. (Tinker & Thornton, 1992, p. 155)

CHAPTER 1 INTRODUCTION

Immersive games have tremendous potential to fill the gulf between what students do for fun and what they are asked to do in school by combining curriculum materials with compelling storylines (Shute et al., 2010). Specifically, for science education, games also are capable of supporting learners' dynamic process of building and justifying explanations. It has the affordance to reify learners' conceptions and models of the world for sense making (Gobert et al., 2015). Several improvements in learning aspects are attributed to educational video games, such as enhancement of students' engagement and motivation, or the potential to offer authentic learning experiences where students encounter open-ended real problems (Tuzun, Yilmaz-Soylu, Karakus, Inal, & Kizilkaya, 2009). Given the potential of game-based technologies to empower learning, policy leaders such as the National Science Foundation (Borgman et al., 2008), the Department of Education (2010), and the National Research Council (NRC, 2011) established a national agenda for game-based learning research.

Nonetheless, the application of serious games in education is not without barriers. One of the most relevant challenges faced by educators is games lack an assessment infrastructure to optimize the learning potential. An obvious difficulty for assessment in games is to implement the "Question and Answer" structures in it. In games, students are constantly solving problems at a certain pace, which is designed to engage students in a fun, challenging but not frustrating experience (Serrano-Laguna et al., 2012). Well-designed games tend to induce flow, a state in which game players lose track of time and are absorbed in the experience of play (Shute et al., 2010). Explicit introduction of traditional assessments in any form is likely to disrupt flow and break the immersive atmosphere.

In practice, teachers usually rely on external assessment to determine the achievements of the students in game-based learning contexts. External assessment is not part of the game-based environment, which is often realized through debriefing interviews, knowledge maps or causal diagrams, and written exams (multiple-choice questions or essays) (Ifenthaler, Eseryel & Ge, 2012). These assessments are often performed as pre and post test or tasks before or after a specified period of game play and focuses on the outcome. Teachers can only compare the individual learning outcome with previous outcomes, check against other learners or experts (Ifenthaler, Eseryel & Ge, 2012). Therefore, they neglect important changes during the learning process. External assessments treat games as a black box since these assessment results might not give enough information about the possible cause for an incorrect result. Neither can assessment conducted after playing the game involve diagnostic feedback that could help shape the student's progression through the game.

Embedded assessments such as game scoring and performance milestones in games are other popular assessment methods in game-based learning. These forms of assessment focus on targets achieved or obstacles conquered while playing the game. Another metric for game scoring is the time consumed in finishing a specific task (Reese et al., 2012). Some researchers refer to these assessment approaches as embedded assessment since they are part of the game mechanic and do not interrupt the game (Reese et al., 2012). However, these assessments are guided by simple performance criteria: the only relevant thing checked is whether the player achieves sufficient performance milestones within the constraints of the game rules.

From an educational point of view, merely concentration on player performance (game scoring on the performance milestones) is not necessarily meaningful for assessment of learning especially for games with more freedom and towards more open ended nature. Several researchers (VandeWalle, Brown, Cron, & Slocum, 1999; Fisher & Ford, 1998) point out the difference between a performance orientation and a learning orientation: while game play tends to center on performance, linking to an attitude of achieving milestones and score, learning demands opportunities for reflection, self-evaluation, pauses, and even the preparedness to make mistakes. Having completed an educational game successfully does not necessarily mean flourishingly learning.

Westera, Nadolski and Hummel (2014) further argued that the discrepancy between learning and performance in games became larger as games provide more freedom of movement to the learners. They indicated while the learning achievement are likely to coincide with performance gains in well-structured drill-and-practice games (Math games or spelling games), the quality of learning is prone to diverge from the quality of performance when games offer more freedom of movement and autonomy as linked with contextualized problem solving, adventure games, and inquiry-based learning, which are common in game-based science learning. As the games become more complex and behavioral variability among individual gets larger, the less information is known about the learner's process of learning and its effectiveness (Westera, Nadolski & Hummel, 2014). Therefore, the video game, again, is treated as a black box that is assumed to generate pre-established learning outcomes, but it is difficult for the embedded game score driven assessment methods to certify the preciseness of our assumption, and to evaluate the process of learning and its quality.

It is pertinent to build an embedded assessment infrastructure to provide insights into the individuals' behaviors, activities, and efforts (flow and grow processes) that the learners demonstrate while playing the game in order to achieve the games' milestones and learning. These insights can shed light on questions such as whether the learner achieve the milestones in an effective and advised style or reaching them in a thoughtless trial and effortless way without much actual learning gains (Westera, Nadolski & Hummel, 2014). Thus, tracking students' interaction trail with the game can reveal the actual in-game student behaviors and trajectories. Exploiting players' traces in games and analyzing how these students' behaviors, wanderings, and psychometric states correlate with learning gains could be beneficial for assessing the process of playing and studying, identifying bottlenecks in game play, and for constructing automatic models to personalize learning. Focusing on the learning process and linking it with the learning outcome instead of solely centering on the outcome is also a general call from the learning science and science education community (Winne and Nesbit, 2010; Gobert, Pedro, Raziuddin & Baker, 2013).

The filed studying the collecting, using and analyzing this kind of learning traces is known as Learning Analytics. It is the "…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" (LAK, 2011). These trace data log all students' interactions within the games, and if fully leveraged, can provide rich assessments of students' learning process and results in games. Embedded assessment from the lens of learning analytics has multiple benefits compared to the external assessment and game scoring approaches (Figure 1). First, it assesses learners while playing the game, providing detailed insights into the underlying learning process. Second, learning analytics can infer higher level knowledge and psychological traits that help us better understand specific behavior and the final learning outcomes. Third, the assessment results of learning analytics have the potential to enable immediate feedback and inform intervention design. Last but not the least, learning analytics for assessment is a subtle process and does not affect any other forms of assessment. In fact, learning analytics mechanism from the author's perspective is not deemed as a replacement for other assessing approaches but as a complement to the existing methods. This is in favor of the statement that embedded and process-oriented assessment must always include multiple measurement procedures which aim to provide reliable and valid evaluation results (Ifenthaler, Eseryel & Ge, 2012).

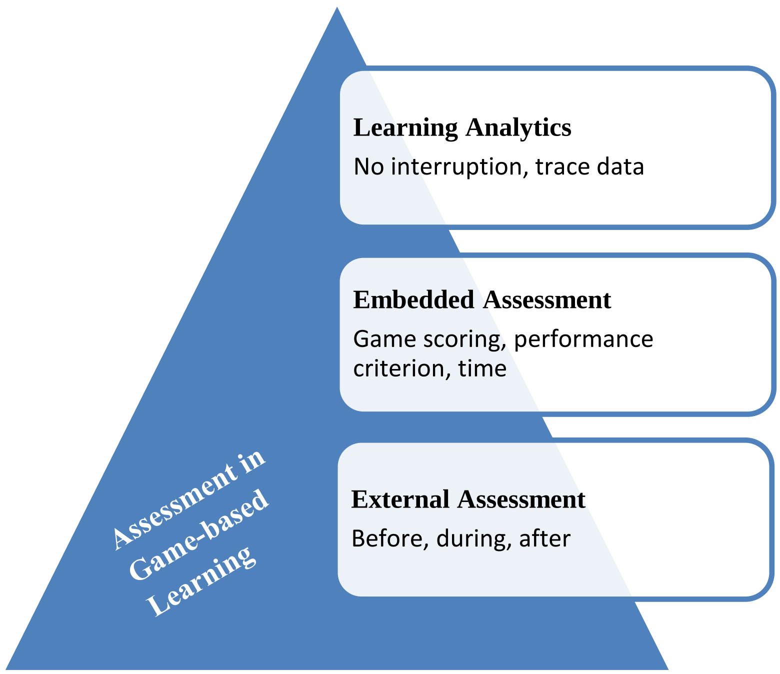

Figure 1. Assessment methods in game-based learning

Several studies have explored the use of electronic data traces for embedded assessment in games from three perspectives. The first one is mainly conceptual work to examine how to collect trace data in games and the conceptual architecture of learning analytics in game-based learning (e.g., Serrano-Laguna, Torrente, Moreno-Ger, & Fernandez-Manjon, 2012; Nelson, Erlandson, & Denham, 2011). The second one relies on simple calculation of the log data using descriptive statistics and visualizing some metrics instead of inferring higher level competence or learning (e.g. Loh, 2006; Kim & Chuang, 2012). It requires teachers' to further deduct the student learning status or results, which might be difficult for some of them who lack the understanding of the mechanics underlying the metrics calculation.

The third line of work using trace data for assessment in games is mainly derived from psychometrics. It favors evidence-centered design (ECD from GlassLab, Mislevy et al., 2014), a rigorous and detailed cognitive framework to guide the assessment in games (e.g. Shute et al, 2010) and then combined with a latent variable modeling technique (e.g. Bayesian Network) for learner assessment in games. The employment of ECD is to ensure the assessment validity and reliability requirements by accomplishing a logically coherent, evidence-based argument between the domain being assessed, assessment task design and interpretation (Mislevy & Haertel, 2006). Sometimes, such assessment approach is also referred as "stealth assessments", measuring learning using tasks embedded within the gameplay itself to "support learning, maintain flow, and remove (or seriously reduce test anxiety), while not sacrificing validity and reliability" (Shute et al., 2010, p. 10). Some of the examples can refer to SimCityEDU from GlassLab (Mislevey et al., 2014), Newton's Playground (Shute, Ventura, & Kim, 2013); and Surge (Clark et al., 2011).

In ECD, learning measures (or assessment mechanics) must be designed along with the game mechanics. However, there is a constant tension between ECD to assure that gameplay is designed to support and measure meaningful learning, while is still open to significant learning that may take place during gameplay, but is not considered by the designers from the start (Rowe, Asbell-Clarke, & Baker, 2015). There may also be a tension between the most enjoyable game mechanics and the most effective learning and assessment mechanics. Therefore, this purely psychometric assessment, a top-down process, focusing on measurement and testing, may limit the pedagogical design for fun and learning in games. Moreover, even though this theory informed assessment design in games, the assessment result of the probabilistic framework is simple possibilities for students' cognitive traits. It is difficult to interpret these probabilities as actionable intelligence to improve students' learning.

In this dissertation, I propose a combination of top-down (theory-driven) and bottom-up (data-driven) approach for learning analytics in game-based learning. In order to inform assessment of learning in games to provide actionable intelligence for teachers and learners, I operationalize theoretical principles to guide learning analytics design (Xing, Guo, Petakovic & Goggins, 2015; Gobert et al., 2015). This theory informed analytics can be principles for distilling, parsing and aggregating the large volumes of log data while students are interacting with the environment. I then employ machine learning algorithms to explore the theory-informed space to build data-driven metrics that can be interpreted in relation to pedagogical theories and instructionally meaningful ways for actionable insights.

Unlike ECD to require a simultaneous consideration on game mechanics and assessment mechanics, the theory in this work is mainly for guidance and interpretation in dealing with data analytics. This theory informed analytics seeks to remain as open as possible to emergent evidence of learning in games while pursuing the logical coherence of theoretical principles. In other words, rather than placing any constraints on the game design, the proposed method allows the assessment mechanism to emerge from the theoretical guided observations of gameplay.

Specifically, this work first theorizes game-based science learning as interactive model based learning, a theory originating from model-based learning and interactivity in multimedia learning. Then trace data, questionnaire data, and learning outcomes data are identified and operationalized to represent these theoretical principles. Finally, machine learning algorithms are designed, adapted, and applied on the gathered data to build automatic assessment models for students' playing and learning process in games as well as linking with the learning outcomes to provide actionable intelligence. In sum, this learning analytics system is expected to enable automated, authentic, and scalable assessment in game-based science learning.

Learning analytics is a cycle of development (Clow, 2012). This work will focus on the first cycle of this development with specific attention on presenting the types of traces to log to facilitate assessment, evaluation of these trace data, and developing metrics and models for embedded assessment. Further evaluation and development are needed with more testing subjects and accumulating more data to train the proposed models. This dissertation is structured as follows: Chapter 2 presents the literature of the previous work on assessment methods in game-based environment, as well as the background of learning analytics and theories applied in this dissertation. Chapter 3 provides the research context and the detailed methodology; Chapter 4 shows results of the developed metrics and models; Chapter 5 discusses the results, summarizes this study, pointing out limitations and future research directions.