Using in situ Data Collection to Improve the Impact and Return of Investment of Game-Based Learning

Christian Sebastian Loh, Ph.D.

Virtual Environment Lab (V-Lab)
Southern Illinois University Carbondale, IL.,
United States of America

csloh@siu.edu

Abstract

Today's economic situation requires all organizations to be more efficient in their business dealings, to reduce cost and increase bottom line. Chief Learning Officers (CLO) of companies must ensure that any investment in learning products will improve the impact and return of investment (ROI) for their companies. Game-based learning (GBL) applications claim they can improve learning, but thus far, have failed to produce concrete, empirical evidence of doing so. As academics and pundits continue to argue the merits of GBL, every passing day constitutes a lost opportunity for the GBL industry. One problem would be the lack of new and appropriate assessment methodology that could showcase the effectiveness of the learning technology, as well as convince stakeholders that GBL works.

This paper describes a new assessment methodology designed specifically for GBL by collecting player-generated data using telemetry in an *in situ* manner (i.e., directly within the GBL environment itself). This approach makes for better assessment than current data-collection methodologies that take place outside the game environment because the data collected is attributed directly to the players' learning performance. Data visualization techniques used in this methodology will further help stakeholders translate the data collected into meaningful information and actionable intelligence.

Keywords: assessment, game-based learning, Information Trails©, *in situ* data collection, telemetry

Introduction

Games-based learning (GBL) is increasingly being considered as the next new learning platform for online training (from military, to corporate training and higher education). Unfortunately, since games and simulations are not all created equal, they can be inconsistent in what they claim to do. For example, does a game really teach, or does it merely provide an exploratory environment where learning 'could' happen? Does a game have an integrated infrastructure to track players' decisions and evaluate them, or does it merely generate a log file and leave it to the trainers to figure out what the data means? Bear in mind that different approaches of assessment could translate to less, or more work for the trainers or instructors. In addition, not all trainers are experts in data analysis, and thus, have the know-hows to make sense of the hidden information.

Use traveling as an analogy, a traveller who has paid an agent to arrange for a trip has every right to expect to arrive at the right destination. Similarly, a manager who invested in game-based learning applications for their organizations would want to know, "Did our trainees arrive at the learning destinations as claimed by the game application or company?"

A learning activity that is without assessment is informal and comparable to the endeavour of hobbyists, at best. As Michael and Chen (2005) suggested, assessment is what differentiates serious games from entertainment games. Without an appropriate means of measuring success in GBL, a self-proclaimed serious game is no better than its entertainment counterparts. Moving forward, it is expected that performance assessment will become an integrated feature in GBL applications due to the demands of the market. While cost-benefit has frequently been mentioned in relation to technology investment (Bates, 1999), a recent survey of Chief Executive Officers (CEOs) of Fortune 500s companies, once again confirmed that Return of Investment (ROI) and impact of the learning products (technology) invested ranked far higher than the process of learning in their organizations (Phillips and Phillips, 2010). This means that stakeholders really cared very little about how craftily GBLs are made, unless these products can be shown to generate impacts and ROIs for their organizations!

The Problem of Assessment with Games

For the most part of the 20th century, assessment has always been conducted within a physical face-to-face environment. However, the advent of advanced computing technology and Internet (as disruptive technologies) has, not only introduced new affordances for learning, but also great changes to the known assessment process. Large international corporations (including the U.S. military) were able to mitigate training cost with virtual environment because trainers and trainees can now be co-located without the need for physical transportation. While the new digital online environment made it extremely easy to conduct distributed training, it also made it very difficult to collect data for assessment.

As there is no safe way to put a probe into the mind of a learner (regardless of the learning environment) to directly sample the amount of learning that occurs, educators must rely on external indicators for performance assessment and evaluation. Within a physical face-to-face environment, a trainer can observe the learners' classroom behaviors (e.g., yawning, or on-task discussion) and document them as evidence of participation and as assessment of the learning that occurred (Harrington, Meisels, McMahon, Dichtelmiller, & Jablon, 1997).

Unfortunately, as there is yet to be any meaningful ways to directly observe or document learners' behaviors in GBL, educators have thus far been prevented from using these well-tested classroom techniques to assess the learning of their students. What is the digital equivalence of an engaged or bored facial expression? Are they paying attention to the learning materials or are they simply skipping over the tedious readings only to focus on the gaming aspect of the GBL? Currently known assessment methodologies are simply not adequate in addressing some of these issues, much less providing answers for these questions.

Formative and Summative Assessments

Despite what many have in mind, assessment of learning is not a monolithic process, but one that evolves with the learning process. An appropriate assessment methodology needs to take into consideration how to best measure the amount of learning that occurred and document the learning that has taken place.

Educators speak of two different types of assessment. The After Action Reports (AAR) used by the military and year-end school examinations are examples of 'summative assessment.' Its purpose is to measure learners' understanding, retention, or mastery of the subject *after* a course of instruction has been completed. Formative assessment, on the other hand, is designed to measure the amount of learning that is taking place while the course of instruction is still *in-process*; it may take the forms of silent observations, feedbacks, peer reviews, and may occur multiple times over a course of instruction.

Formative assessment is by far the more valuable of the two assessments because the information produces usable knowledge for the instructors, enabling them to make improvement on a personal as well as at the systemic level. As learning and content management systems (such as Blackboard and Moodle) add automated feedback and grading as standard features in their online technology-support learning environments, today's learners will come to expect the same to be available in GBL (Bloxham & Boyd, 2007; Parkin & Thorpe, 2009).

GBLs that capture simple metrics of 'number of kills' or 'number of missions accomplished' to constitute high scores and 'Leaders' Board' may claim that they are 'doing assessment.' However, data captured in this manner is far too simplistic to produce the level of impact sought after by stakeholders, which is obtainable only through in-depth analysis of play-learners generated data. Such analysis will eventually produce what Rao (2003) referred to as '*actionable intelligence*.' As more stakeholders make demands for this kind of high level actionable intelligence, it will hopefully 'motivate' the GBL industry to incorporate real assessment into GBL applications as a standard feature in the future.

A Much Needed Collaboration

It is foreseeable that as the GBL industry continues to grow, it will increasingly be expected to: (1) collect data for formative assessment, (2) turn the data collected into *actionable intelligence* via information visualization processes, and (3) ease the calculation of impact and ROI for the CLO – who reports to the CEOs of the organizations. These innovative features are not currently available in GBLs, and will furthermore, require specialized expertise in data handling and information visualization. The quickest way to achieve this is for game companies to collaborate with game assessment experts.

Unfortunately, insiders' information from the game industry indicated that game industry has thus far been reluctant to work with the academics (Hopson, 2006). This explains the current status as to why there are hardly any GBLs that provide stakeholders with the performance assessment indices they seek. Speaking objectively, without the means to properly assess learning, GBLs are just mere games. Persistence in the current course of action will eventually lead to the repeat of *edutainment* history (van Eck, 2006), where products fail, both in education and entertainment. Once the market became saturated with GBL that produces no verifiable results, stakeholders might conclude this as another overhyped technology: a conclusion we would like to see.

Performance Gap

Performance Gap Analysis suggested that the lack in assessment for GBL (see Figure 1) to be due to a lack in: 1) resources, 2) motivation and 3) knowledge. In this case, since the game industry is such a highly creative and resourceful enterprise, no lack of resources is expected. As explained in the previous section, GBL developers will continue to experience a lack of motivation until the assessment of GBL can be translated into revenue for the game companies. Unfortunately, instructional designers understand that unlike a lack of knowledge, a lack in motivation is not something that can be overcome through training.

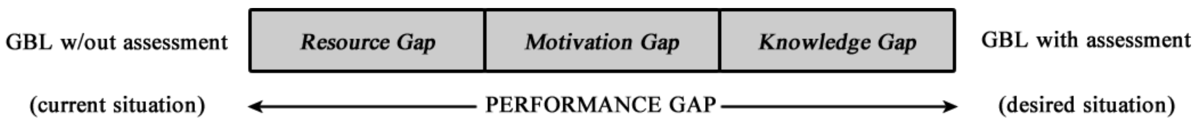

Figure 1 – Performance Gap Analysis

On the other hand, training is an appropriate bridge to overcome the *knowledge gap*. Game developers who are not familiar with conducting learning assessment would do well to work with a game assessment researcher who understands the demands from both sides. Not only are game assessment researchers far more familiar with designing assessment according to the needs of the training/education sectors, they are inevitably gamers (even modders), thus placing themselves in a unique position to bridge this gap.

Ex situ vs. *In Situ* Data Collection

Although game assessment is still a relatively new area of research, current methodologies fall into two general approaches:

1) *ex situ* data collection: GBL is seen as a black box, therefore data collection has to be conducted 'outside the box,' and

2) *in situ* data collection: GBL is seen as a white/open box, one could therefore take advantage of the software (engine) to collect data 'inside the box.' This will be explained in details in the following sections.

Game-Based Learning as a 'Black Box'

A good example where GBL is likened to an impenetrable black box is the category of 'miracle drugs.' These are drugs that apparently improve a patient's condition, but where researchers are yet to fully understand how the drug really 'works.' Drug companies which sell the miracle drugs thus based their claims solely upon observations of external indicators manifested by patients (who received the intervention). Most commonly, pre-test and post-test are administered – before and after an intervention, respectively, to try and reduce the 'noise' in the data. The idea being, if the workings of the 'drug' is unknowable (i.e., black box), then by observing the knowable (conditions of the patients) before and after the drug administration, the difference between post-test and pre-test must equal the effects of the drug.

While this approach is sufficient for the drug industry and traditional classrooms, it carries problems for GBL. Not only will no one ever know what really happen inside the black box (in this case, GBL), the positive changes could be due to a number of extraneous factors that are accounted for, or worse, achieved through players '*gaming the system.*' [For a thorough explanation on 'gaming the system,' see Baker, Corbett, Koedinger, & Roll (2006); Baker, Corbett, Roll, & Koedinger, (2008).]

This approach is rather popular among educator-researchers because it is an assessment processes that has been used for many years in the traditional classroom settings, to measure the effects of new education technologies, ranging from television, to learning machine, to individualized computer-assisted instruction.

Game-Based Learning as an 'Open Box'

People who have had computer programming background, however, see GBL to be a vastly different entity. Instead of a black box, they see GBL as just another software application, and hence, a 'white/open box.' If this is an open box, then it should be possible to 'peek inside' the box to observe the workings of the gears within, and perhaps to measure the performance of individual components within the system (with specialized test-pens).

As users interact with software applications (in this case, GBL), new user-data are being generated on-the-fly. Because these data (internal 'variables') are generated through the actions and behaviors of users of GBL, they are valid evidence – similar to the heartbeats and respiratory rates generated on-the-fly by users of 'exercise machines.'

These variables, found in the GBL system memory, can then be selected, filtered, and stored in a database for later retrieval and analysis. There would be no external 'noise' in this case because data collection occurred within a 'closed environment.' Figure 2 depicts the differences between the two approaches of data collection with GBL.

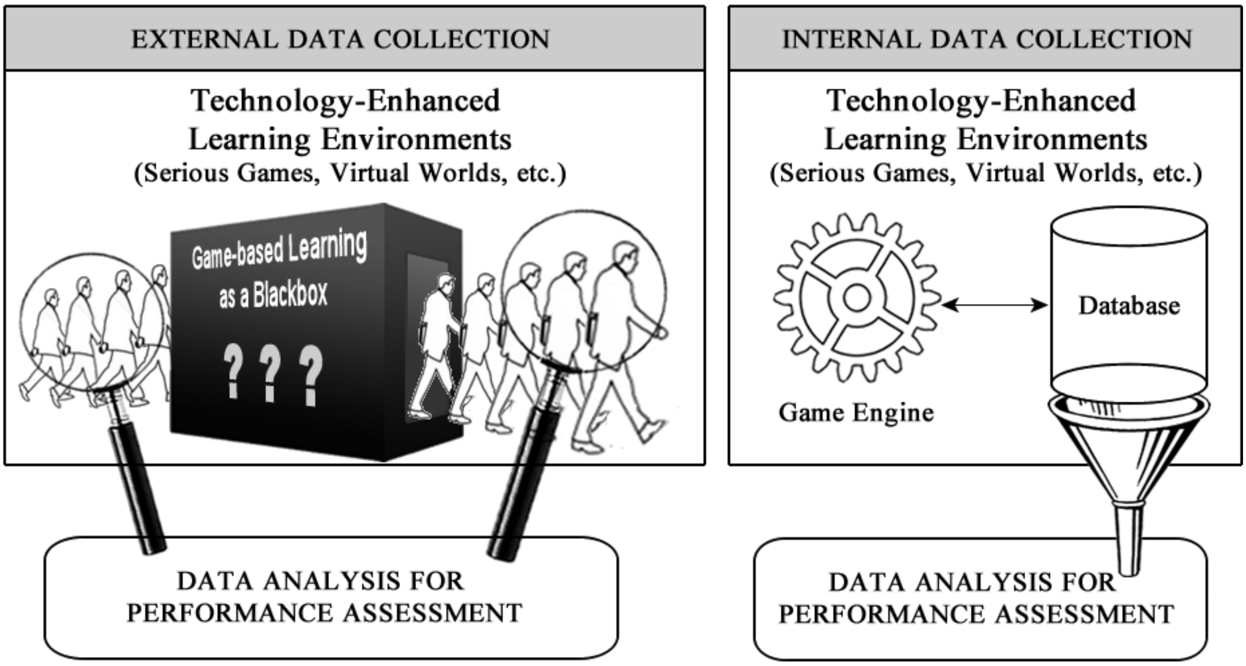

Figure 2 – *ex situ* (left) vs. *in situ* (right) Data Collection Approach with Game-Based Learning

Increasing Impacts of GBL Using an *ad hoc* Assessment System

From a research stancepoint, understanding that GBLs are not black boxes opens up new ways to conduct assessment. In addition, these 'internal variables' (i.e., player-generated data) may even be sampled multiple times *during* a game play session for formative assessment. Researchers no longer need to wait until the game is fully completed before they can begin analyze the data for performance assessment. Since some GBLs can last as long as 20-40 hours, an *ad hoc* (as opposed to a *post hoc*) sampling process will also help organizations obtain assessment reports much earlier in the process and not to wait until the end of the 20-40 hours.

An *ad hoc,* just-in-time/on-demand sampling will also help to generate *actionable intelligence* faster, enabling the trainers to be informed of the learning progress of the trainees earlier, and give them a chance to prescribed remediation in a timely manner. Since 'catching' human errors early can also help save re-training cost in organziations, it would directly impact ROI, something stakeholders look for.

Take pilot training for example, a recent finding revealed severe flaws in the flight simulation systems used by airlines to train their pilot, resulted in several airline accidents (Levine, 2010). How much money would the airline companies have saved if they were the ones who first discovered the flaws and took the necessary corrective measures at an earlier time?

The following sections described an *ad hoc* performance assessment approach, known as *Information Trails*© that was designed to take advantage of the software engine in GBL for internal data collection (Loh, 2007). However, because considerable expertise are required to make sense of the massive amount of data collected, a software reporting system (work-in-progres) has also been developed to visualize the data into useful information (Loh & Li, 2010).

In situ Data Collection Using Information Trails©

No matter the environment (virtual or physical), a learner's actions and behaviors are ultimately determined by his/her decision making process. Using a GBL scenario, if a particular path leads to certain death (say, careless mistake leading to an explosion), players must find alternative means/routes to solve the given problem. This leads many game companies to believe GBL is just another way to conduct exploratory-learning, and by allowing players to find the 'correct combination of steps' using a 'trial-and-error' approach, they have done a good job. However, it is far more important to make sure the players did not find 'making mistakes' to be fun, but that they truly understand the factor, or combination of factors, that could result in a present danger.

Because *Information Trails©* allow trainers to 'see' the repetitive actions performed by players, the trainers can now determine if players' mistakes are repeated because of ignorance (meaning remediation is needed), or just for fun (probably an undesirable action). The emerging pattern of behaviors can then be inferred to reveal the person's decision making and reasoning processes. Since a decision is the product of a person's knowledge schema, the effectiveness of a user's actions – speed, accuracy and strategy – within the information ecology can then be expressed as a function of the users' understandings of the learning problems (what they know), as well as their problem solving skills (what they can do).

Conceptually, *Information Trails©* comprised of a series of strategically placed and agent-traceable objectives (events) within any information ecology (including GBL and virtual worlds). Much like traffic cameras, 'event recorders' are positioned at strategic locations (nodes) to capture user-actions during key events. Once captured, the users' actions may then be analyzed at any time using any appropriate method to reconstruct the decision-making process of the providers. By analyzing the logics of the decision-making process, actionable intelligence (Rao, 2003) can then be produced, either to supplement a 'beneficial or profitable' decision, or to correct a 'risky' decision that may lead to financial loss.

In practice, the *Information Trails©* approach employed a telemetric process to collect the data generated by players. Telemetry is a technological process (used in many industry, including medical and wildlife research), in which the objects of interest were tagged with technological devices that allowed remote tracking, and the data collected by these devices were compiled into metrics and remotely sent back to the researcher for recording and analysis. It should be noted that the *Information Trails©* framework does not, by design, capture all available information indiscriminately. The author firmly believes that choosing what data to capture is just as important as what data *not* to capture.

In order for *Information Trails©* to work with existing *GBL*, an *event listener* (software) must be made available. Using the event listener, events in the GBL that are deemed important for the learning process (i.e., *actionable objectives*) can then be tracked. These *player-generated action data* can then be stored in a *database server*, and be transformed into human-friend reports in real-time via information visualization techniques. A corresponding reporting tool (detailed in section 4.1) is currently under development to turn raw data into useful information, or *actionable intelligence*.

The following figure (Figure 3) shows the relationships among game engine, event listener, external database server, actionable learning and game objectives, and the *ad hoc* reporting system/tool. It should be obvious that without the assessment components, a standalone GBL engine will only produce more games that cannot be assessed.

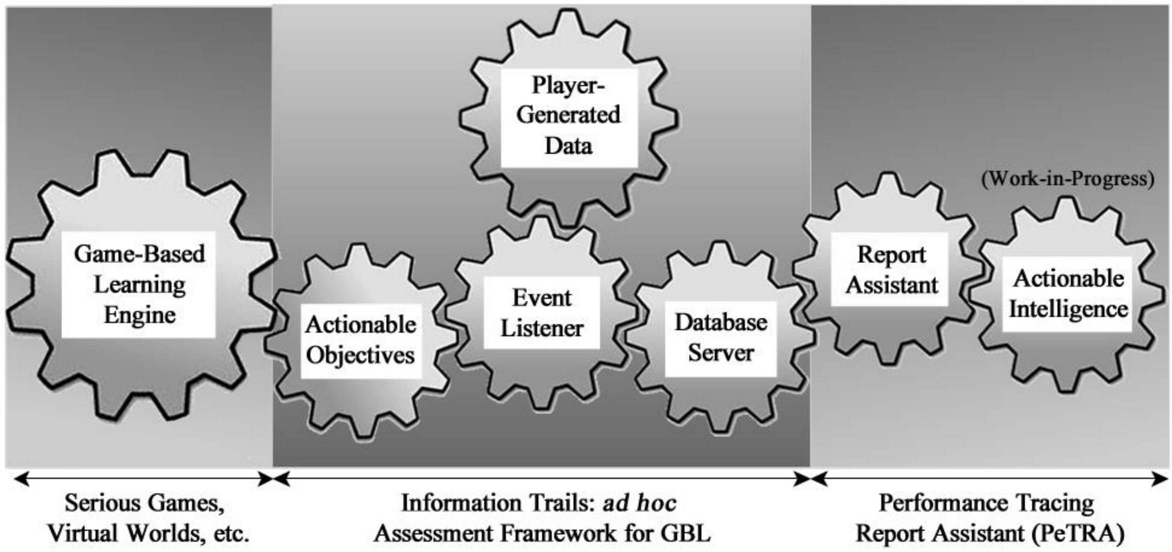

Figure 3 – Relationships among various components of GBL with formative assessment capability

Performance Tracing Report Assistant© (PeTRA)

If *Information Trails©* is a framework to capture learner-generated data within GBL, then *PeTRA©* is the display showcase of those data as useful information. In the hands of a capable CLO, these information could easily be translated into actionable intelligence. Data visualization is an extremely important step in this case because it provides trainers and CLOs a window into the GBL and allows them to 'observe' the training while it is still in-progress. In addition, since most individuals are not trained to handle vast amount of data, PeTRA helped to make sense of the data via a human-friendly report interface; in essence, it is a specialized reporting tool for *Information Trails©*.

The availability of the reporting tool means there is no need for organziations to hire in-house data analysts to do the job: a cost saving measures. Since *PeTRA©* is created using *Adobe Flash Builder*, it can also be viewed over the Internet and on mobile devices, thus allowing real-time distributed debriefing between trainers and trainees. These are all features that are not possible with a paper-based/printed report.

Information Visualization

Data analysis packages (such as SPSS, MathLab) are not sufficient for this type of analysis because they do not contained information visualization processes. Because each GBL is created for a different audience and purpose, the visualization and report will also differ accordingly. This is why it is important to begin planning for *Information Trails©* and assessment report as early as possible in the game development process, and not to retrofit the assessment component into an existing GBL.

Two examples with information visualizations are presented below to illustrate the need for this important feature. The first example (Figure 4) shows the path traversed in a single GBL session by one player. Without additional (visual) information, it would be difficult for trainers to understand why the player traversed the GBL environment as such. With more information included (such as a bird's eye view of the area map), trainers would have more than sufficient visual cues to understand the reasons behind the player's movements, actions and behaviors. Event and position markers taken at regular intervals further helped to 'connect the dots.' In the example shown below, the line and the circular dots depicts the path traversed and interaction points generated by a trainee from the Beginning to the End (blue square dots) of the training session.

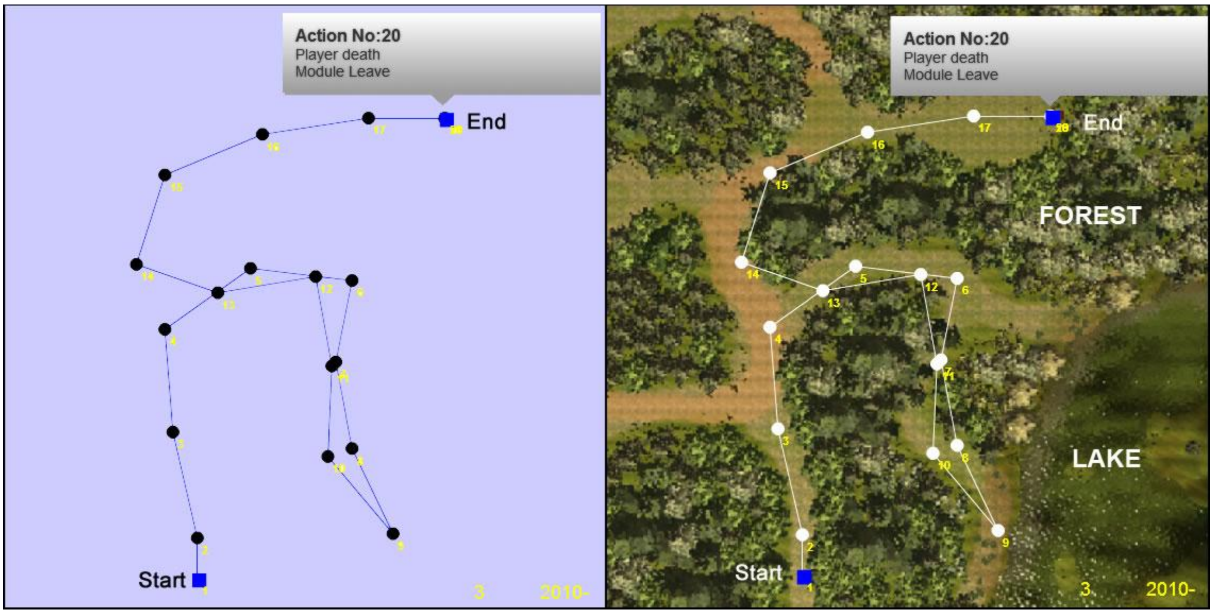

Figure 4 – Path traversed by player during GBL with (right) and without (left) additional visual cues.

Information visualizing technique is also useful in performance comparison of learners. The following example (Figure 5) reveals the deviations found when one expert's performance was compared against that of a novice, and an intermediate learner (i.e., in between a novice and an expert).

Using expert behaviors as the baseline, the infograph revealed three distinct anomalic patterns indicating unnecessary actions/tasks performed by the novice and intermediate learners. More specifically, Anomaly No.1 and No.3 are extraneous tasks performed by the novice, and intermediate learners, respectively. Whereas tasks found in the Anomaly No. 2 zone are performed by both the novice and the intermediate learners.

By narrowing down to the task and objective levels, trainers would be able to analyze the actions of novice trainees in greater details and provide them with appropriate and individualized advise to improve their learning performance. Further analysis of the tasks found in the Zone 1 and 3 should further explain the apparent mutual exclusivity, and the reasons the novice (Zone 1) and intermediate (Zone 3) learners did what they do.

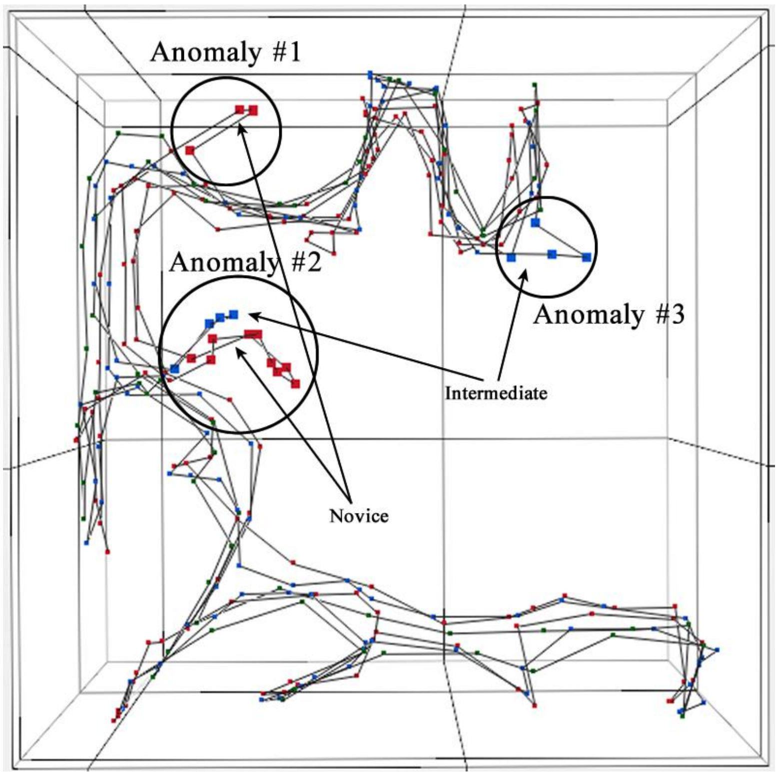

Figure 5 – Comparison of performance among expert, intermediate, and novice players.

Conclusions

Information Trails© and *PeTRA*© brought a number of advantages to GBL. Instead of having to wait 20-40 hours for the GBL to be completed, trainers could now debrief trainees earlier and at a more frequent interval. Since *PeTRA*© can be used both for *ad hoc* and *post hoc* reporting, those who require *post hoc* debriefing will benefit from the 'instant replay' function of the reporting tool. The mobile device compatibility will ease deployment within training organizations and at the same time, allowing distributed debreifing.

The real-time reporting capbility of *PeTRA*© means that trainees will be informed of the impact of their actions in GBL faster, so that they can modify their actions and behaviors accordingly to yield maximum performance. Man-made errors committed during training can be detected earlier before they became greater problem, thus saving training organizations valuable time and money.

The use of GBL has the potentials to revolutionize the way people learn and how organziations train their workforce. However, without appropriate assessment components (e.g., the data collection infrastructure, reporting tools), stakeholders of learning organziations will have a difficult time distinguishing GBL from (entertainment-based) games. Learning processes are less important to stakeholders when compared to measurable impacts and ROI of the learning products they invested in. Integration of *Information Trails*© and *PeTRA*© into a GBL will not only add assessment capability, but also the compelling evidence that the application works as claimed.

New technology often requires new assessment methodology to showcase its effectiveness and to provide stakeholders with the evidence they need to make the investment. Linda G. Roberts, ex-Director of Education Technology to the U.S. Department of Education, once said, "I believed that researchers could improve the design and collection of data. Just as new technology created new opportunities for learning, it created ways to invent new tools for research and evaluation, particularly ways to track and monitor what, how, and when learning occurred" (2003, p. viii).

The lack of appropriate assessment methodologies tailored for GBL has not bode well for the learning products. Already, pundits and skeptics have begun criticizing GBL as just one more educational technology that is "useless", "ineffective," and showing "no significant difference" in improving learning and eductaion (c.f. Cuban, 2001, and Clark, 2007). If the field of GBL assessment is to grow, it will require the collaboration between game assessment rsearchers and GBL companies to produce the kind of products that will yield high impacts and generate ROI for learning organziations.

Game assessment research is still very much in its infancy, with the first edited book project being prepared for publication by Ifenthaler, Eseryel & Ge (2012) at the moment. In the next few years, as 'learning analytics' gain greater importance among the learning organizations, developers will have no choice but to incorporate telemetry into their games to facilitate the collection and assessment of play-user generated (learning) data. However, this only constitutes half the answer because it is unfair to expect game developers to know which metrics to use for learning performance assessment. These questions ought to be answered by educators, trainers, and CLOs of companies.

Data visualization is yet another area that will see tremendous growth in the coming future because it is the only means by which stakeholders can make sense of the massive amount of data collected into actionable intelligence. Instead of reinventing the assessment wheel at every turn, researchers and game developers should collaborate to solve common problems for the advancement of the field.

References

Baker, R. S. J. d, Corbett, A. T., Koedinger, K. R., & Roll, I. (2006). Generalizing detection of gaming the system across a tutoring curriculum. Paper presented at the 8th International Conference on Intelligent Tutoring Systems. Jhongli, Taiwan.

Baker, R. S. J. d, Corbett, A. T., Roll, I., & Koedinger, K. R. (2008). Developing a generalizable detector of when students game the system. *User Modeling and User-Adapted Interaction*, *18*(3), 287-314.

Bates, A., W. (1999). Managing technological change, strategies for college and university leaders. San Francisco, CA: Jossey-Bass Publishers.

Bloxham, S., & Boyd, P. (2007). *Developing effective assessment in higher education*. Berkshire, UK: Open University Press.

Clark, R. E. (2007, May-June). Point of view: Learning from serious games? *Educational Technology, 47*, 56-59.

Cuban, L. (2001). *Oversold and underused: Computers in the classroom*. Cambridge, MA: Harvard University Press.

Harrington, H. L., Meisels, S. J., McMahon, P., Dichtelmiller, M. L., & Jablon, J. R. (1997). *Observing, documenting and assessing learning*. Ann Arbor, MI: Rebus, Inc.

Ifenthaler, D., Eseryel, D., & Ge, X. (Eds.). (2011). *Assessment in Game-based Learning: Foundations, Innovations, and Perspectives*. New York, NY: Springer, New York.

Levin, A. (2010, August 30). Simulator training flaws tied to airline crashes. *USA Today*. Retrieved from http://travel.usatoday.com/flights/2010-08-31-1Acockpits31_ST_N.htm.

Loh, C. S. (2007). Designing Online Games Assessment as Information Trails. In D. Gibson, C. Aldrich, & M. Prensky (Eds.), *Games and Simulations in Online Learning: Research and Development Frameworks* (pp. 323-348). Hershey, PA: Information Science Publishing.

Loh, C. S., & Li, H. I. (2010). Reducing re-training cost through on-demand, *ad hoc* assessment. Paper presented at the MODSIM World Conference and Expo 2010 – 21st Century Decision-Making: The Art of Modeling & Simulation (Oct 13-15, 2010). Hampton, VA: MODSIM World Conference & Exposition.

Parkin, H., & Thorpe, L. (2009). Exploring student experiences of e-learning: A phenomenographic approach. Paper presented at Bera Annual Conference (September 2-5, 2009). Manchester, UK: University of Manchester.

Phillips, J. J., & Phillips, P. P. (2010). How executives view learning metrics. *Chief Learning Officer*, *9*(12), 82-84. Retrieved May 20, 2011, from http://clomedia.com/articles/view/how-executives-view-learning-metrics.

Rao, R. (2003). From unstructured data to actionable intelligence. *IT Professional*, *5*(6), 29-35. doi: 10.1109/MITP.2003.1254966.

Robert, L. G. (2003). Forewords. In G. D. Haertel & B. Means (Eds.), *Evaluating educational technology: Effective research designs for improving learning* (p. 290). New York, NY: Teachers College Press.

Citation: Loh, C. S. (Sep 2011). Using in situ data collection to improve the impact and return of investment of game-based learning. In proceedings of the 61st International Council for Educational Media (ICEM) and the XIII International Symposium on Computers in Education (SIIE) Joint Conference (ICEM-SIIE 2011). Aveiro, Portugal: ICEM-SIIE.