



Using players' Gameplay Action-Decision Profiles to prescribe training: Reducing training costs with Serious Games Analytics

Christian Loh and I-Hung Li

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

August 31, 2018

Using Players' Gameplay Action-Decision Profiles to Prescribe Training

Reducing Training Costs with Serious Games Analytics

Abstract— Player data from Serious Games can be used to produce Serious Games Analytics, which can in turn, be used to measure, assess, and improvement performance. These insights can also be used to support decision-making by Chief Learning Officers in 'prescribing' training – e.g., diagnosing who should receive training, how to design serious games for optimal training, and what content should be included or withheld. Data-driven training prescription can help learning organizations save money by mitigating unnecessary training sessions and identifying what kind of training fits which individual to eventually reduce training costs.

We traced the player data and calculated the similarity between their course of actions (COAs) compared to that of the expert's *in situ* serious-game training environments. We found the combined metrics of Cosine similarity and Maximum Similarity Index to be useful in identifying players' Gameplay Action-Decision (GAD) profiles. Insights to using GAD profiles as a diagnostic to measure training performance and prescribe training are offered. (Abstract)

Keywords—*Serious Games Analytics, Cosine similarity, expertise levels, performance improvement, reducing training cost.*

I. INTRODUCTION

The first commission to convert an entertainment game into a serious game purposed for training originated in the U.S. Marine Corps [1]. Commandant Krulak's idea was to co-opt computer-based gaming technology into *tools* to improve thinking and decision making skills at a time when training resources became increasingly limited; thus, moving the training sessions from the real world (live training) into a virtual one. Nearly 20 years later, that seed grew not only into a legitimate field of research but also a thriving industry of development.

Serious games currently available are varied in their purposes, with about 90% of them are educative *message broadcasters* aimed at information dissemination [2], [3]. The remainder are *tools* that are capable of bringing about 'sustained instruction and practice (given or received) in an art, profession, occupation, or procedure, with a view to proficiency in it' (definition of *training*, Oxford English Dictionary), or 'relative permanent change in user-behaviors as a result of interaction with the environment' (definition of *learning* [4, p.30]). Serious games designed with training and learning in mind can be suitable for acquiring skills not easily taught in classrooms, including cognitive processes such as strategic and analytical thinking, planning and execution, problem solving, decision-making, and adapting to rapid change [5]. This model of serious games is also suitable for

extension beyond the traditional classroom for rehabilitation [6], patient choice education [7], and expertise training [8].

In order to evaluate if competency is gained during training, performance assessment must take place. For serious games and other virtual environments (VE) for training, software telemetry must be put in place to track and monitor user behaviors *in situ* such environments [9]. Because serious games will need to contend with the performance of users, telemetric approaches must be designed to have been designed to capture user-interactions with *both* the training and gaming elements as behavioral data.

It is the purpose of *Game Analytics* to 'maximize the value of player data' [10]. Similarly, *Serious Games Analytics* is a concerted effort to maximize that value using new and improved 'methodologies for performance measurement, assessment, and improvement' [11] – to better the training processes using serious games. New research to predict expertise performance is also available [12]–[14] and can be useful for the calculation of the success rates of training.

With more serious games deployed, it is becoming easier to collect (big) data on play-user behaviors. What remains lacking is a set of established (business) analytics methods to systematically obtain meaningful insights and use the newfound knowledge to predict performance, or to prescribe training. By '*prescribe training*', we mean the ability to identify *who*, *what*, and *when* train should occur, or *not* occur – decisions that can directly impact training costs in organizations.

Since training can be a very costly endeavor from an organizational point of view, even the U.S. military has to find ways to reduce training overheads [1], [15]. It should not surprise anyone that Chief Learning Officers with good business acumen will actively seek to reduce training cost to improve their bottom lines and ensure better returns. One quarter of the Global Fortune 500 companies have already adopted serious games for training [16]. A more streamlined approach to designing serious games with assessment and guidelines to prescribe training will grow the market even more. A performance assessment component with analytics to prescribe training will be particularly useful in the market segment for military, medical, and surgical training [17], for starters.

II. MOTIVATION

The loss of expertise through retirement and turnover is inevitable for any learning organization. Retired experts must be replaced by new workers to ensure the survival of the organization. Inexperienced workers (some of whom will be

complete novices) must be trained to an acceptable level of expertise (i.e., competent, or proficient) before they can be added to the workforce. Further expertise beyond the Proficient level would require time and deliberate practice (see Fig. 1).

In most practical situations, training the new workforce will involve just three levels: novice, competency, and proficiency. Further training to the level of expert and master (i.e., over-training) is not practical because those levels can only be attained through long periods of deliberate practice [18], [19]. As the main purpose of training is to “ensure the acquisition of required knowledge, skills, and abilities” [21, p. 143], the effectiveness of the training program may then be quantified through how many competent (lower limit) to proficient (upper limit) individuals were produced. A competent individual should have acquired basic knowledge, skills, and abilities. Those at the Proficient level should know when to apply those knowledge, skills, and abilities, such as realizing if a mistake has been made, knowing when to self-correct, etc.

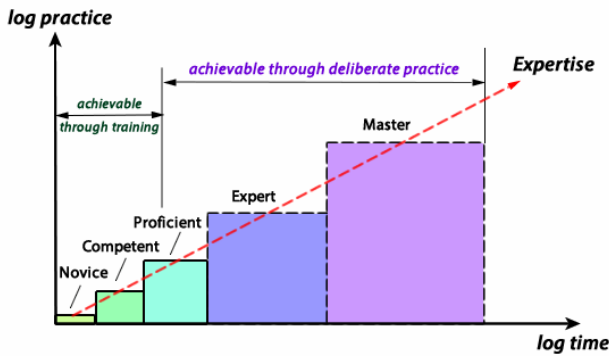


Fig. 1 Effects of training and deliberate practice on levels of expertise.

It is important to differentiate over-training from excessive training. Some amount of over-training, has been shown to lead to automaticity and is necessary in maintaining adequate performance during high-stress situations [21], [22]. Excessive training, on the other hand, is indiscriminate un-prescribed training that is wasteful and expensive. However, under-training can be more expensive than excessive training because it can become a liability – e.g., rookie mistakes turning into lawsuits or insurance claim.

The possibility of flawed training models being incorporated into VE training can be hard to detect [23]. Companies that are using or planning to adopt serious games (or other VEs) for training need to be able to identify *who*, *what*, and *when* to train, or *not* to train – so as to mitigate excessive and under-training, and to identify flawed training before too late. Hence, the need for *training prescription*.

This paper calculate the *Expertise Index* of trainees by comparing their *course of action* with VE training against that of an expert’s baseline, using the concepts and procedures presented by[24]–[26]. In addition to existing methods, we will describe *Cosine* similarity and how it can also be useful for the calculation of similarities between expert and novice COAs. We will then discuss the implications of the *findings* for training prescription (such as who, and who not to train).

III. COMPETENCY AND COURSE OF ACTION

Users’ *competency* can be observed and demonstrated through their chosen *course of action* (COA) during training [19]. This concept has been used to convert users’ COAs in VE training for performance assessment analysis by way of similarity indices [24]–[26].

A. Similarity Indices

For example: most people can agree that the text-string, APPLE, is nearly identical to APPLES, but not so much with ORANGE. But is there a way to formally quantify the degree of (dis)similarity among the strings? We can use the *Similarity Index* – a statistical method to standardize the quantification of (dis)similarities between texts and documents. The value of the *Similarity Index* (ranges from 0 to 1) provides an indication of the similarity between two strings: a value of 1 means the strings are identical, and a value of 0 would mean they are completely dissimilar.

Using the 3x3 map shown in Fig. 2 as an example, we can denote the navigational COA of a user (blue) as StringA (ABEDGHI), and the COA of an expert (red) as StringB (ABEFI). More complex games with lots of events and markers may require different grid sizes to better capture the COAs for analysis, see [14].

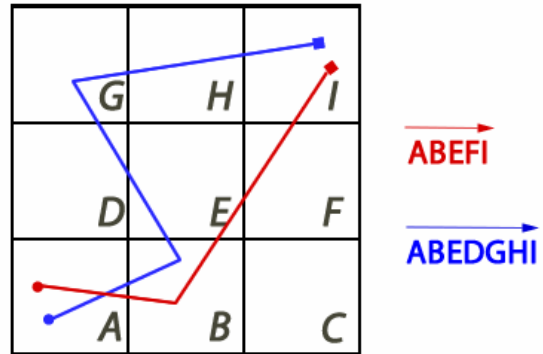


Fig. 2 Navigational *course of actions* (COA) for User (blue) and Expert (red).

By comparing the COAs of the entire user corpus to a (fixed) expert’s COA, the obtained similarity indices will reveal how closely each user’s performance is as compared to that of the expert. As shown in Fig. 3, higher indices (towards the right) would place the users nearer to the level of proficient, whereas lower indices would indicate that the users are more like novices. Competent users are found around the middle region.



Fig. 3. Similarity index and trainable levels of expertise.

B. Cosine Similarity

The similarity between two strings can be quantified using string metrics [28], such as Cosine similarity or Jaccard coefficient. Cosine similarity has become quite popular in

information retrieval and data mining for a number of complex tasks, including document clustering [29], facial verification [30], PageRanking (an algorithm used in Google’s Search engine [31]), and others.

$$\cos(A, B) = A \cdot B / (\|A\| \times \|B\|) \quad (1)$$

Since the formula for Cosine similarity (Eq. 1) was designed for the computation of similarity between two vectors, any strings intended for comparison must first be converted into vectors before the formula can be applied. The total vector space in the example above (i.e., StringA and StringB) can be defined as {A, B, C, D, E, F, G, H, I}. The strings can then be converted into vectors as shown in Table 1.

TABLE I. STRINGS TO VECTORS CONVERSION

Vector Space	A	B	C	D	E	F	G	H	I
StringA (ABEDGHI)	1	1	0	1	1	0	1	1	1
StringB (ABEFI)	1	1	0	0	1	1	0	0	1

The dot product (nominator) is $A \cdot B = 4$, while the product of the magnitudes (denominator) is $\|A\| \times \|B\| = \sqrt{7} \times \sqrt{5}$. For our example, the Cosine similarity between StringA and StringB is therefore: $4 / (\sqrt{7} \times \sqrt{5}) = 0.676$.

In the next section, we will describe the methodology and how we collect user data *in situ* the (in-house designed) serious game for this study.

IV. METHODOLOGY

The Federation of American Scientists listed the following features as important for training using serious games [5]: (a) having clear goals, (b) learning through repeatable tasks to build mastery, (c) monitoring of learners’ progress, (d) encouraging increased time-on-task through motivation, and (e) matching difficulty level to learners’ level of mastery – i.e., adaptive learning. As this study is about determining the level of mastery of users based on their decision making processes *in situ* the virtual environment of a serious game, only the first four features (i.e., a to d) were included.

We used the Unity3D game engine to create a large one-level Maze for our study. This Maze contained only one Exit and the goal of the game is to search and locate the Exit (to escape it) as fast as possible. We have designed two *critical paths* [32] for this game: routeA has no obstacle and is slightly longer than routeB, which has a locked door and a pressure plate on the floor for triggering the lock mechanism. Players must step on the pressure plate repeatedly (for 9 times) before the door could be unlock and access to the Exit given. The excessive number of steps were intended to separate players who are merely curious about the pressure plate from those who would not give up easily (i.e., being *persistent*). Learners lacking in volition would probably give up on the pressure plate without thinking too deeply about it, translating to a lower Cosine similarity index for routeB.

Both critical paths would eventually lead to the Exit room containing a portal (i.e., magic circle). Stepping into the circle

will cause the player to leave the Maze and be shown an on-screen notice. They would be shown how long it took them (in number of seconds) to find the Exit, and be asked if they would like to try for a faster time by re-entering the Maze or end the game, by clicking on a ‘YES,’ or ‘QUIT’ button, respectively.

The study was approved by the Human Subject Review Board of a large mid-Western university and a total of 16 students volunteered for the study. The students could terminate the gameplay at any time, but would be rewarded with 2, 5, or 7 points for a class assignment, if they could ‘escape’ the Maze in less than or equal to (\leq) 100, 50 or 38 seconds, respectively. All 16 players’ performance were compared against a single Expert, whose records for routeA and routeB were 38 second and 34 seconds, respectively.

V. FINDINGS AND DISCUSSIONS

We use R [33] for all the statistical calculations, and the *stringdist* package [28] to compute the Cosine similarity in this study. We also set the confidence level for *t*-Test at 99% (i.e., α -level = .99). Given that there were two critical paths in the study, it would give rise to two Expert routes. As such, we computed the Cosine similarity twice for each player – i.e., once each for the Expert’s COA of routeA, and routeB.

We plotted the two Cosine similarity indices for all the players and found some interesting patterns. Based on players’ gameplay actions and how they go about making decisions (to solve problems) *in situ* serious games, we came up with a *Gameplay Action–Decision (GAD)* profile for each player. Players’ GAD profiles can be roughly divided into three bins (A, B, and C). We shall refer to them as *Fulfillers (binA)*, *Explorers (binB)*, and *Quitters (binC)* in the next sections. (The terms are not meant to be pejorative but to accurately describe players’ in-game behaviors.)

A. Gameplay Action–Decision (GAD) Profiles

We called the six players in *binA*, **Fulfillers**. The players’ interests are limited to completing the job given to them and nothing more. All 6 players in this category appeared single-minded in that once they have discovered a workable route (be it routeA or routeB), they will keep working on it until the intended outcome has been achieved (Fig. 4). Fulfillers did not seem to question if better alternatives exist. Although some players ended up with high scores, others eventually ‘gave up’ after some time. The amount of time they were willing to spend working on the task seemed dependent on a pre-determined quota that constituted *their part* of the unspoken agreement. This quota appeared to differ from person to person, which explains why some individuals press on to reach better scores.

The four players in *binB* are **Explorers**, as they did not stop at just one solution but were more willing to continue searching for potentially better alternatives. Their performance is characterized by crisscross(es) of COAs as they switch from route to route. Fig. 5 shows one player’s MSI with the first crisscross (B→A) between round 9-10, and the second (A→B) between rounds 12-13. Unlike the Fulfillers, the high score in round 8 did not stop the Explorers from continuing to search for *potentially* better alternative route. By potentially better, we

mean the alternative(s) may or may not lead to better results. Since the game is novel, it is entirely possible for the Explorers to find no better alternative than their first route. It would appear these players are interested enough in the learning tasks to take the extra initiative. They were able to self-evaluate their findings to see if the new route helped to increase performance and will only stop when they are fully satisfied with the outcome – a trait also found in *perfectionists*.

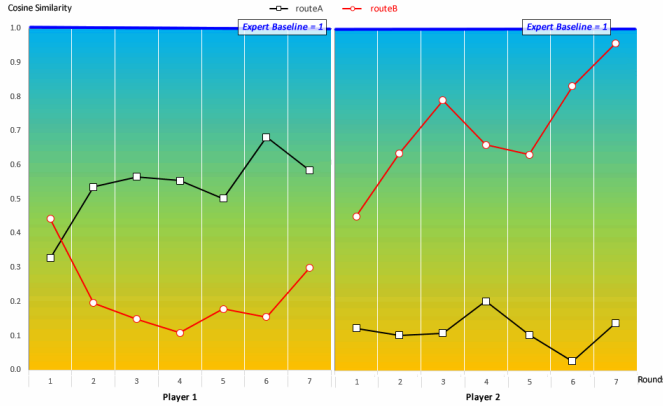


Fig. 4 *Fulfillers* discover a working solution and stay with it (Player 1 found routeA, and Player 2, routeB).

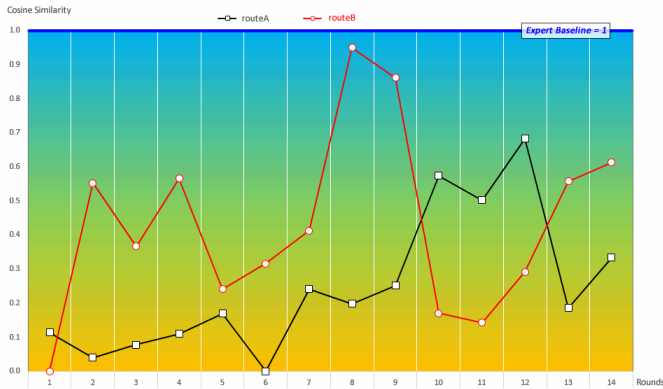


Fig. 5 *Explorers* seek alternatives that can potentially better performance, their play graph shows crisscrosses of strategies or routes.

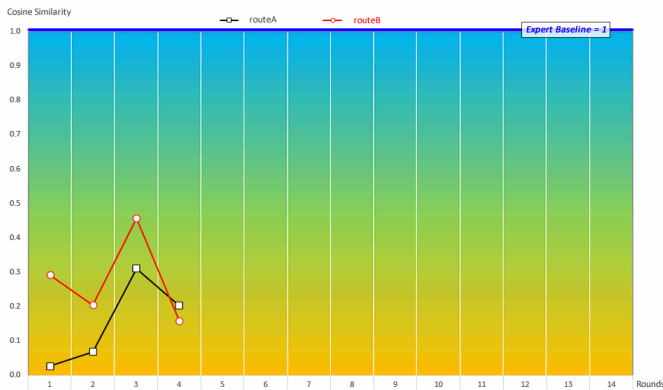


Fig. 6 *Quitters* give up early and tend to remain as novices despite training.

Lastly, we refer to the six players in *binC* as *Quitters*. Their behaviors are characterized by low scores, limited number of trials, and early abandonment (i.e., quitting) with incomplete rounds (Fig. 6). These players appeared to be either disinterested in the training (content?) or simply lacking the motivation to keep trying. As a result, they tended to give up too quickly (in our study, three players dropped out after just 1 round) and resigned to being novices, despite the opportunities to improve their performance through training. Generally, they are having more of a motivational issue than lacking in the ability to learn or improve.

Literature in Instructional Design [34] speaks of the existence of a *performance gap* (Fig. 7) that is created when there is a lack in (a) knowledge, (b) resource, and/or (c) motivation. Research has shown that only the gap of *knowledge* is bridgeable through training. This means that *Quitters* are not likely to respond to training – be it serious games or other virtual environments, until the underlying motivational issues has been dealt with. Thus, for training prescription, it makes more sense to send the *Quitters* to team-building exercises (to raise their morale) than to training courses.



Fig. 7 Performance Gap.

B. Visualizing Performance as Serious Games Analytics

Since there are two critical paths in this game, there are also two Expert routes – resulting in two Cosine similarity indices per player. This means that we will need a new method to ascertain the players’ performance when multiple experts’ routes are present.

Fortunately, the *Maximum Similarity Index* (MSI) method [25] is applicable in this case and can be adopted for use in the next part of the calculation. The MSI approach records the *highest* attained similarity index to compensate for the presence of multiple experts and is especially useful for calculating the achievement of Explorers. Without MSI, the gameplay actions of these players – crisscrossing from one route to the other, will result in an oscillating Cosine similarity curve that is difficult to interpret.

We ranked the MSI of all 16 players (from low to high) and found a large jump between Player 17 and 7 (Fig. 8). To the left of the performance jump are the six *Quitters* (Player 9, 15, 10, 18, 8, and 17), and to the right are the remainder players (*Fulfillers* plus *Explorers*, or non-*Quitters*: Player 7, 19, 4, 16, 5, 22, 13, 20, 6, 2, and 14). The highest MSI scored by the non-*Quitter* group is 0.959 – highly similar to that of the Expert baseline of 1.

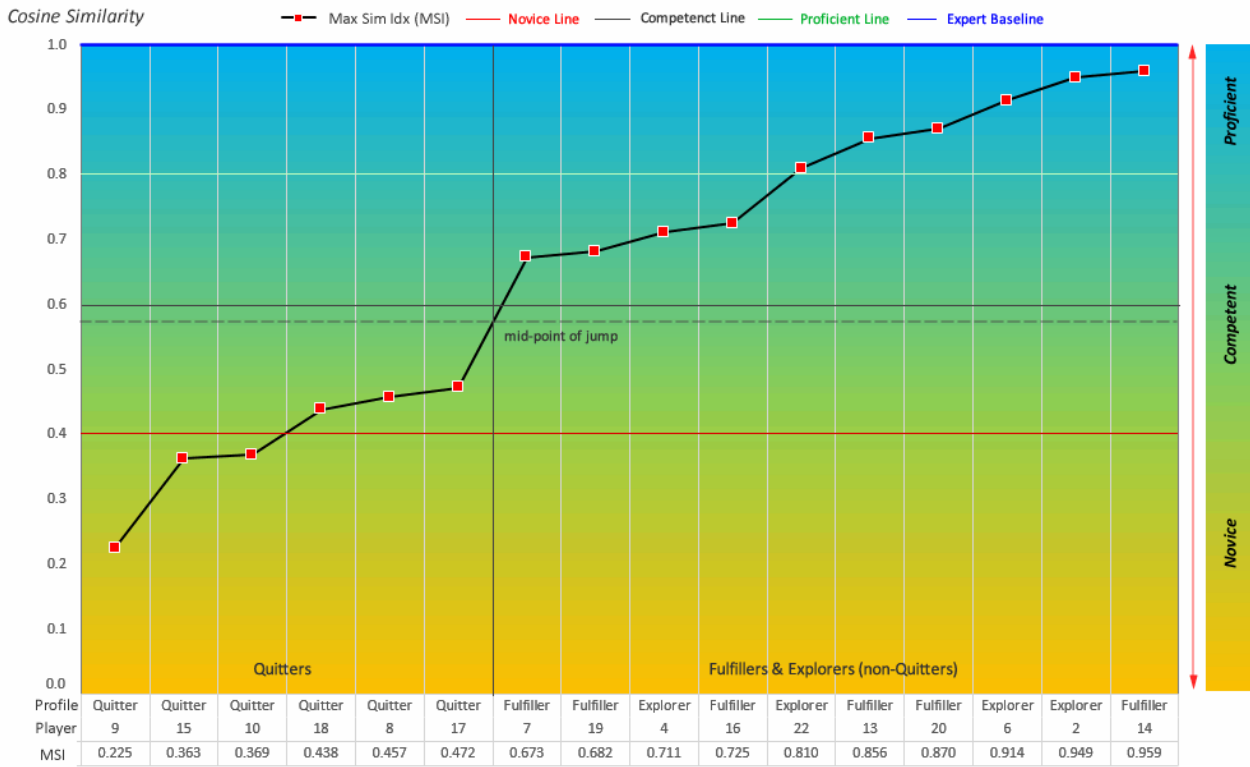


Fig. 7 Players ranked by Maximum Similarity Index.

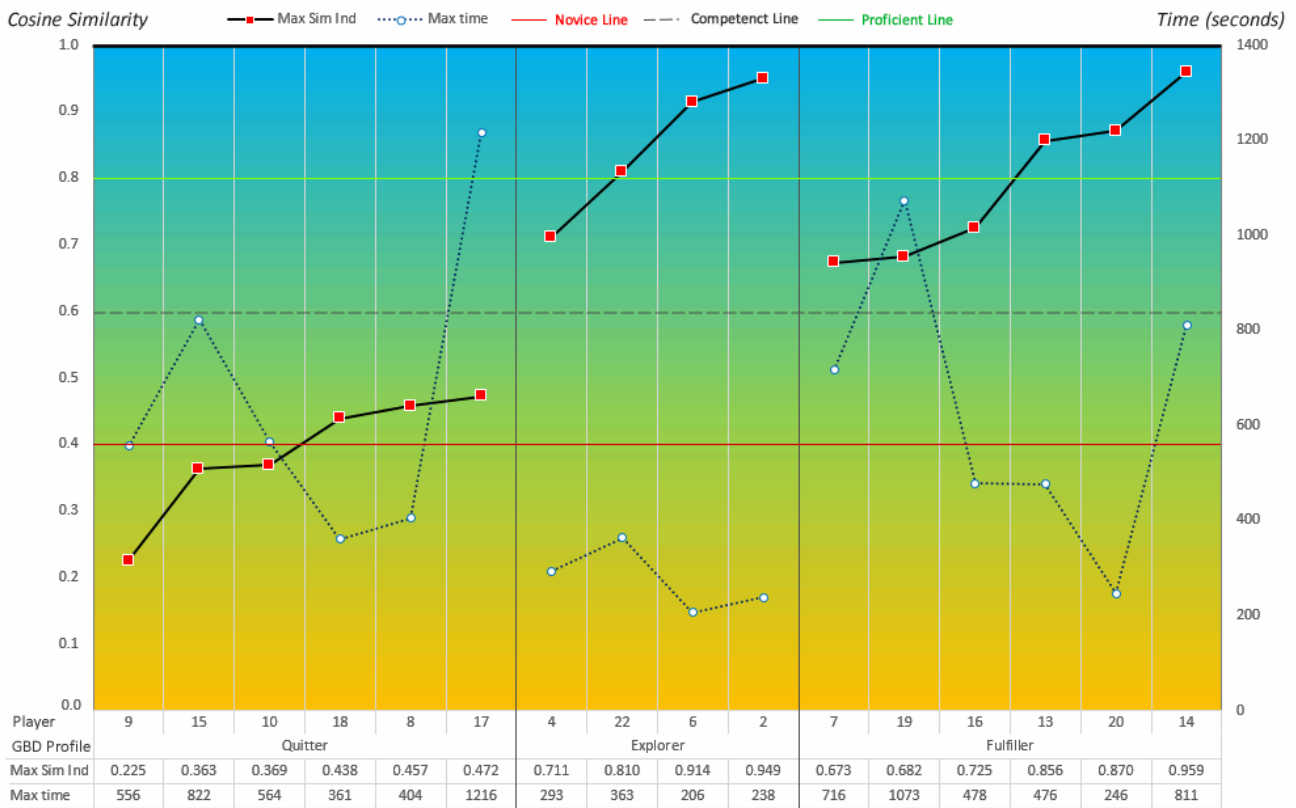


Fig. 8 Maximum Similarity Index plotted against Players' Gameplay Action- Decision Profiles.

C. Gameplay Action-Decision Profiles

We learned that Explorers would seek alternative(s) in the hope that it may be better than the solution they hold. Fulfillers, on the other hand, are less likely to look for alternative if they have a working solution at hand. But how can we maximize player data value with this insight? More importantly, is there any difference between the two profiles in terms of performance? If differences do exist, then should we train individuals to become Fulfillers or Explorers? To answer these questions, we tried plotting the MSIs against the players' Gameplay Action-Decision profiles: i.e., Fulfillers, Explorers, Quitters. We think the answer is quite surprising and may redefine how we do training in the future.

Fig. 8 shows the graph where MSI is ranked and plotted against the Players' Gameplay Action-Decision profiles as the primary axes, and time of completion (in seconds) as the secondary axes. It should be immediately apparent that the performance of Quitters ($M = 0.399$, $SD = 0.068$) is rather poor, while the performance of Fulfillers ($M = 0.794$, $SD = 0.117$) and Explorers ($M = 0.846$, $SD = 0.108$) are much higher and similar. Boxplot of MSIs against the three profiles were shown in Fig. 9.

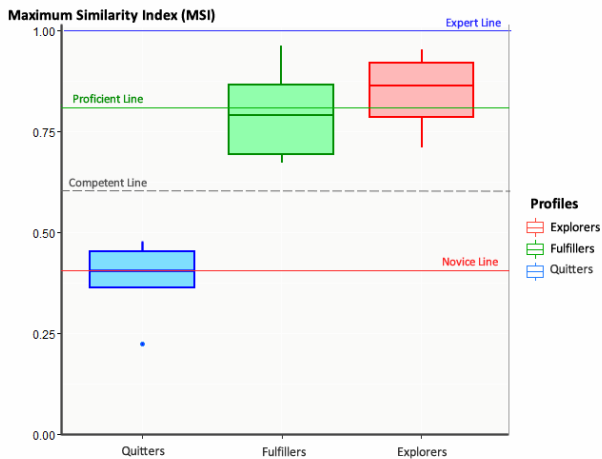


Fig. 9 Boxplot of MSIs vs. Gameplay Action-Decision Profiles.

We conducted a one-way ANOVA to confirm the differences among the mean of the three categories are statistically significant. Further testing with Tukey's HSD found statistically significant differences between the performance levels between Explore-Quitters ($p < 0.0001$) and Fulfiller-Quitters ($p < 0.0001$), but not for Fulfillers-Explorers ($p = 0.805$). In summary, the Quitters did not seem to benefit much from the training, whereas the Fulfillers and Explorers did.

D. Total Time Spent

There was a surprising correlation between total time spent on the training (serious game) and the players' Gameplay Action-Decision profiles. Despite spending less time in the training, the Explorers achieved a similar performance to that of the Fulfillers. It would appear that spending a little time

searching for alternative(s) does pay off. Not only was the performance level of Explorers comparable to that of the Fulfillers, they were able to complete the training under much shorter timeframe. Since completing the training in a shorter time translates to a reduction in training costs, the Explorer is the profile to be modeled after!

The best player in the Fulfiller group (Player 14) achieved a MSI of 0.959 after spending 811 total seconds in the training. The second best player from the same group, Player 20 spent only 246 total seconds to achieve a MSI score of 0.870. By comparison, Player 20 performed much better than Player 14 because the time to expertise ratio is better.

At this time, the scoring of MSI is being calculated independent of the total time spent in game. The original purpose of MSI was to evaluate the decision-making strategy (i.e., COAs) of players in serious games as analytics. This new finding (correlations between players' GAD profiles and total time spent) seemed to indicate that there could be value in combining the two metrics into one for performance evaluation.

E. Expertise Line

Expertise lines can be a convenient decision tool for Chief Learning Officers to visualize the performance of trainees at a glance. Since the jump in performance (between Quitters and non-Quitters) occurred in a range between .5 (50%) and .7 (70%) of the Cosine Similarity index, the approximate midpoint (~60%) seemed a good place for the Competent Line. For the remainder expertise, a ± 2 (20%) separation seemed reasonable. Thus, we placed the Novice Line at $(.6 - .2 = .4)$ and the Proficient Line at $(.6 + .2 = .8)$. The Expert (base)line is always at 1.

However, based on our experience in serious games research, the placement of these expertise lines (Novice, Competent, and Proficient) are highly dependent on the games (and how they affect player performance). Therefore, our recommendation is to use a continuum to denote the changes in expertise level – such as the heat-map visualization shown as the colored background in the graph – rather than to affix them to some numerical values.

F. Future Research

Readers familiar with decision-making styles in educational psychology literature may notice much overlaps between *Fulfillers* and *Satisficer*, as well as *Explorers* and *Maximizers* [38], [39]. Future research can determine if they are subsets of one another, or are in fact, interchangeable terms.

Future research can also look into combining MSI with the total time spent in-game to create a single score for performance ranking purposes. Replicating this study with the new metric may also yield new insights about Fulfillers and Explorers. At the very least, it gives the industry a new metric to work with and allows stakeholders to calculate returns.

We also lack a way to measure the initial performance of the players. Once that data is available, we can better calculate performance differences (or net improvement) by subtracting the initial performance score from the MSI, so as to obtain a

net performance increase. As the players' COAs become increasingly similar to that of expert COAs due to training, the players are growing in competency to become more and more like the expert.

VI. CONCLUSIONS

A. Prescribing Training with Serious Games Analytics

We already know that players' gameplay data can be processed for values [37], [38]. The Federation of American Scientists has suggested that these data can also be used to create *adaptive learning* in serious games [5]. Findings from this study has yielded new insights on how to transform players' gameplay action-decisions into profiles. More research will be needed to provide instructional and serious games designers with better understandings on how players' Gameplay Action-Decision profiles contribute to performance outcomes in serious games. By understanding how players make decisions (during interactive training) in serious games, we can better design the training paths and game events to (re)channel the flow of training.

People joined an organization with various prior experience and learning habits – not all are desirable. The ability to predict performance and know when to support or when *not* to support training is important. The aim of *training prescription* is to diagnose how people learn *in situ* technology-based environment, predict patterns, and prescribe (recommend) training based on the outcome of the diagnosis. Unlike *adaptive learning*, which is aimed at supporting how people learn, *prescriptive training* can support, withhold, or even reform old learning habits. Examples of reformative learning include remediation/re-training, habit reformation, and rehabilitation.

Players' Gameplay Action-Decision (GAD) profiles can be used as Serious Games Analytics to prescribe training in the following cases, by helping stakeholders identify:

- *Who* to receive and who *Not* to receive training – Besides identifying Explorers and Fulfillers to participate in training, early identification of Quitters with intervention can help to minimize workplace problems and increase overall morale. The money saved from not sending the Quitters to training can then be rechanneled to send them to team-building exercises (aimed at raising motivation and morale). The money can also be used on gamification, or incentives, to raise external motivation. Unlike training, such actions may actually have a chance in improving the performance of the Quitters.
- *What* types of contents to be offered – Fulfiller would benefit from guided learning contents that is more linear, whereas Explorers would be more suited for discovery-based learning that has an open approach.
- *How* training should be designed – Because Fulfillers are so single-minded in their approach, they may not be open to explore new routes on their own accord. This means the solutions must be somewhat obvious or Fulfillers may miss them altogether. Training programs

designed for Fulfillers should, therefore, have either a linear design, or as few learning paths as possible to ensure thorough familiarization by these players. On the other hand, serious games designed with Explorers in mind should have many alternatives to possibly encourage further exploration. However, the game may become too complicated and costly to produce. In general, workforce training could benefit more from a Fulfiller-style (focused) design, whereas managerial training could benefit from an open, Explorer-style, design.

- *Supportive or Corrective Training* – instead of reducing the number of alternatives for Fulfillers (as suggested above), serious game designers can purposely require Fulfillers to consider all available alternatives, or reward Explorers who make decisions based on first available working solution. Such training may go against the natural 'grain' of the players' behavioral makeup, but can be useful for targeted training (or reprogramming) to help them overcome old habits, and acquire new knowledge, skills, and abilities.
- *Stress Training* – “the primary purpose of stress training is to prepare the individuals to maintain effective performance in a high-stress environment” [21, p.143]. While most training programs try to reduce cognitive loads and not tax trainees to the point of fatigue to ensure some level of performance, stress training (though unpleasant) can be just as effective, if not more, in raising performance. Work situations are often stressful (especially in the military), requiring players to complete training tasks under time pressure [26], [39], limited resources, noisy environments, and even sleep deprivation could help prepare individuals to commit fewer errors during emergencies. Stress training is also the way to go for disaster preparations (using serious game) because such training enhances familiarity, builds confidence in the trainees [36]. This approach has been shown to reduce stress on 'game day' for sport players [41]. Judicial use of stress training can also reduce costly mistakes made by workers and learners because they are less likely to be caught by surprise or act out of panic.

Not matter how motivating 'games' can be, serious games can never be that pill to cure those who are unmotivated to learn. We can, however, turn players' data from serious games into new insights that will take us one step closer to *prescriptive training* – identify who, what, and how best to connect with training more effectively to reduce training costs.

REFERENCES

- [1] C. Krulak, "Military thinking and decision making exercises," Washington, DC, 1500.55, 1997.
- [2] C. S. Loh, Y. Sheng, and D. Ifenthaler, "Serious games analytics: Theoretical framework," in *Serious Games Analytics: Methodologies for Performance Measurement, Assessment, and Improvement*, C. S. Loh, Y. Sheng, and D. Ifenthaler, Eds. Switzerland: Springer International Publishing, 2015, pp. 3–29.

- [3] J. Alvarez, D. Djaouti, O. Rampnoux, and V. Alvarez, "Serious games market: Some key figures (from 1950's to 2000's)," 2011. [Online]. Available: http://serious.gameclassification.com/files/articles/sgc_report_03-11.pdf.
- [4] Albert J. Kearney, *Understanding Applied Behavior Analysis*, 2nd ed. Philadelphia, PA: Jessica Kingsley Publishers, 2015.
- [5] Foundation of American Scientists, "Summit on Educational Games: Harnessing the power of video games for learning," Washington, DC, 2006.
- [6] D. J. Cornforth, A. Koenig, R. Riener, K. August, A. H. Khandoker, C. Karmakar, M. Palaniswami, and H. F. Jelinek, "The role of serious games in robot exoskeleton-assisted rehabilitation of stroke patients," in *Serious Games Analytics: Methodologies for Performance Measurement, Assessment, and Improvement*, C. S. Loh, Y. Sheng, and D. Ifenthaler, Eds. Switzerland: Springer International Publishing, 2015, pp. 233–254.
- [7] R. P. Mihail, N. Jacobs, J. Goldsmith, and K. Lohr, "Using visual analytics to inform rheumatoid arthritis patient choices," in *Serious Games Analytics: Methodologies for Performance Measurement, Assessment, and Improvement*, C. S. Loh, Y. Sheng, and D. Ifenthaler, Eds. Switzerland: Springer International Publishing, 2015, pp. 211–231.
- [8] C. McLeroy, "History of military gaming," *Soldiers Magazine*, Arlington, VA, pp. 4–6, Sep-2008.
- [9] G. Zoeller, "Game development telemetry in production," in *Game analytics: Maximizing the value of player data*, M. Seif El-Nasr, A. Drachen, and A. Canossa, Eds. London, UK: Springer London, 2013, pp. 111–136.
- [10] M. Seif El-Nasr, A. Drachen, and A. Canossa, Eds., *Game analytics: Maximizing the value of player data*. London, UK.: Springer London, 2013.
- [11] C. S. Loh, Y. Sheng, and D. Ifenthaler, Eds., *Serious Games Analytics: Methodologies for Performance Measurement, Assessment, and Improvement*. Switzerland: Springer International Publishing, 2015.
- [12] C. S. Loh and I.-H. Li, "Predicting the Competency Improvement for Serious Games Analytics: Action-sequences, Game Grids, PLS DA and JMP," in *Discovery Summit*, 2015.
- [13] Y.-E. Liu, T. Mandel, E. Butler, E. Andersen, E. O'Rourke, E. Brunskill, and Zoran Popović, "Predicting player moves in an educational game: A hybrid approach," in *Proceedings of the 6th International Conference on Educational Data Mining (EDM 2013)*, 2013.
- [14] C. S. Loh, Y. Sheng, and I.-H. Li, "Predicting expert-novice performance as Serious Games Analytics with objective-oriented and navigational action sequences," *Comput. Human Behav.*, vol. 49, pp. 147–155, 2015.
- [15] C. S. Loh, "Improving the impact and return of investment of game-based learning," *Int. J. Virtual Pers. Learn. Environ.*, vol. 4, no. 1, pp. 1–15, 2012.
- [16] BankersLab, "A smart guide to serious gaming," 2013. [Online]. Available: <http://bankerslab.com/blogposts/a-smart-guide-to-serious-gaming-part-1/>. [Accessed: 22-Dec-2014].
- [17] Ambient Insight, "2013-2018 North America mobile edugame market," Monroe, WA, 2013.
- [18] K. A. Ericsson, M. J. Prietula, and E. T. Cokely, "The making of an expert," *Harv. Bus. Rev.*, vol. 85, no. 7–8, pp. 114–21, 193, 2007.
- [19] K. A. Ericsson, N. Charness, P. J. Feltovich, and R. R. Hoffman, Eds., *The Cambridge handbook of expertise and expert performance*. New York, NY: Cambridge University Press, 2006.
- [20] J. E. Driskell, E. Salas, and J. H. Johnston, "Decision making and performance under stress," in *Military Life: The Psychology of Serving in Peace and Combat*, 1st ed., T. W. Britt, C. A. Castro, and A. B. Adler, Eds. Westport, CT: Praeger Security International, 2006, pp. 128–154.
- [21] J. E. Driskell, R. P. Willis, and C. Copper, "Effect of overlearning on retention," *J. Appl. Psychol.*, vol. 77, no. 5, pp. 615–622, 1992.
- [22] D. Rohrer, K. Taylor, H. Pashler, J. T. Wixted, and N. J. Cepeda, "The effect of overlearning on long-term retention," *Appl. Cogn. Psychol.*, vol. 19, pp. 361–374, 2005.
- [23] A. Levin, "Simulator training flaws tied to airline crashes," *USA Today*, 30-Aug-2010.
- [24] C. S. Loh and Y. Sheng, "Measuring the (dis-)similarity between expert and novice behaviors as serious games analytics," *Educ. Inf. Technol.*, vol. 20, no. 1, pp. 5–19, 2015.
- [25] C. S. Loh and Y. Sheng, "Maximum Similarity Index (MSI): A metric to differentiate the performance of novices vs. multiple-experts in serious games," *Comput. Hum. Behav.*, vol. 39, pp. 322–330, 2014.
- [26] C. S. Loh and Y. Sheng, "Performance metrics for serious games: Will the (real) expert please step forward?," in *Proceedings of CGAMES 2013 USA*, 2013, pp. 202–206.
- [27] S. E. Dreyfus, "The five-stage model of adult skill acquisition," *Bull. Sci. Technol. Soc.*, vol. 24, no. 3, pp. 177–181, Jun. 2004.
- [28] M. P. J. van der Loo, "The stringdist package for approximate string matching," *R J.*, vol. 6, no. 1, pp. 111–122, 2014.
- [29] A. Huang, "Similarity measures for text document clustering," in *Proceedings of the 6th New Zealand Computer Science Research Student Conference (NZCSRSC 2008)*, 2008, pp. 49–56.
- [30] H. V. Nguyen and L. Bai, "Cosine Similarity Metric Learning for face verification," in *Computer Vision – ACCV 2010*, R. Kimmel, R. Klette, and A. Sugimoto, Eds. Springer Berlin Heidelberg, 2011, pp. 709–720.
- [31] J. Vembunayanan, "Tf-Idf and Cosine Similarity," 2013. [Online]. Available: <https://janav.wordpress.com/2013/10/27/tf-idf-and-cosine-similarity/>. [Accessed: 12-May-2016].
- [32] T. Short, "Level design in procedural generation," *Gamasutra*, 04-Feb-2014.
- [33] R Core Team, "R: A language and environment for statistical computing." R Foundation for Statistical Computing, Vienna, Austria, 2014.
- [34] R. M. Branch, *Instructional Design: The ADDIE Approach*. New York, NY: Springer Science+Business Media, 2010.
- [35] A. M. Parker, W. B. de Bruin, and B. Fischhoff, "Maximizers versus satisficers: Decision-making styles, competence, and outcomes," *Judgm. Decis. Mak.*, vol. 2, no. 6, pp. 342–350, 2007.
- [36] B. Schwartz, A. Ward, J. Monterosso, S. Lyubomirsky, K. White, and D. R. Lehman, "Maximizing versus satisficing: Happiness is a matter of choice," *J. Pers. Soc. Psychol.*, vol. 83, no. 5, pp. 1178–1197, 2002.
- [37] A. Canossa, M. Seif El-Nasr, and A. Drachen, "Benefits of game analytics: Stakeholders, contexts and domains," in *Game analytics: Maximizing the value of player data*, M. Seif El-Nasr, A. Drachen, and A.

- Canossa, Eds. London, UK: Springer London, 2013, pp. 41–52.
- [38] A. De Gloria, F. Bellotti, and R. Berta, “Serious Games for education and training,” *International Journal of Serious Games*, vol. 1, no. 1. 03-Feb-2014.
- [39] M. E. Young, “Sex differences in the inference and perception of causal relations within a video game,” *Front. Psychol.*, vol. 5, p. 926, Jan. 2014.
- [40] J. E. Driskell and J. H. Johnston, “Stress exposure training,” in *Making decisions under stress: Implications for individual and team training*, J. A. Cannon-Bowers and E. Salas, Eds. Washington, DC.: American Psychological Association, 1998, pp. 191–217.
- [41] T. Driskell, S. Sclafani, and J. E. Driskell, “Reducing the effects of game day pressures through Stress Exposure Training,” *J. Sport Psychol. Action*, vol. 5, no. 1, pp. 28–43, 2014.