Serious Games and Analytics for Skill Acquisition and Assessment

Christian Sebastian Loh, Ph.D.
School of Education
Southern Illinois University

Abstract

By definition, serious games are digital games that are created primarily not for entertainment but for serious purposes such as performance assessment, training improvement, promotion of healthcare practices, and social good. The Dreyfus five-stage Skill Acquisition Model has been proven to be a valuable model with extensive use in professional training fields such as medical, nursing, aviation, and teacher training. However, the use of the Skill Acquisition Model in serious games training, assessment, and analytics is still very new. This chapter introduces readers to serious games and their applications in skill acquisition training and performance assessment. When used in conjunction with new serious games analytics tools such as the *Expert Similarity Index* (ESI) and *Maximum Similarity Index* (MSI), Dreyfus model can help in providing the framework for interpreting findings and bolster performance assessment in serious games to support user performance ranking and the identification of skill stages.

INTRODUCTION

Neuroscience research confirmed that “in order for an experience to become a memory, it must cause a physical change in the brain” (Meister & Buffalo, 2017, p. 694) through the growth of new synapses and the weakening of old synaptic connections. Skill acquisition, as a form of learning, also must be established first through the formation of long-term memory. Some long-term memories involve cognitive tasks (i.e., creative thinking, troubleshooting, strategizing, and self-reflection), while others can only be established through practice via fine controls of musculature movements, (i.e., motor memory). The so-called “muscle memory” is a misnomer as the musculature movements, which reside in long-term memory, are very much controlled by the brain. Examples include martial arts, the playing of musical instruments, sports, and games.

Skill Acquisition and Deliberate Practice

From the perspective of psychology, however, skill acquisition is about the development of procedural (non-declarative) memory. As people learn to perform (a series of) activities that constitute tasks, their performance will gradually improve with (repeated) practice. Some may even discover more effective and efficient ways of completing these tasks, leading to improved proficiency and, potentially, expertise. Feedback, self-analyses, and corrective adjustments for mistakes and weaknesses are additional factors that can result in performance improvement. This consciously conducted, iterative ‘repetition-feedback-evaluation-correction’ cycle, otherwise known as *deliberate practice* (Ericsson et al., 2007; Ericsson & Charness, 1994), is key to improving one’s skill. Research from different fields, especially chess playing and the playing of musical instruments, has consistently reported a minimum of 10,000 hours (or 10 years) of deliberate practice to be necessary in developing expertise.

Although researchers are yet unclear about the relationships among deliberate practice, motivation, and cognitive abilities, many believe that the advancement in digital technologies and training methods can help to *expedite the process* of expertise development (Hoffman et al., 2013). Appropriate use of instructional technology for training such as ‘part-task’ training method (Fadde, 2009) –breaking down a task into trainable subparts for targeted training – and serious games (Knight et al., 2010) are all relevant examples. Analytics – using data science to convert user-generated actions during gameplay for actionable insights – is another useful tool that has many applications for skill acquisition, including performance analysis, user ranking, and training prescription.

The Dreyfus Skill Acquisition Model

Besides deliberate practice, the skill acquisition process can also be interpreted through the lens of the Five-Stage Skill Acquisition Model (Dreyfus, 2004). The five stages of the model are Novice, Advanced Beginner, Competent, Proficient, and Expert. At the start of the skill acquisition process, Novices begin learning by following rules that determine actions. However, since Novices are new to the domain, they tend to follow the (context-free) rules indiscriminately or blindly. As they gain more experience with real situations and events, as Advanced Beginners, they begin to form *maxims* that can be applied in future situations beyond the context-free rules. Having learned the importance of situational and context-free features from even more experience, they start to gain Competence and become able to recognize *aspects* (recurrent patterns) in the environment and select the most salient features to improve their performance. This is accomplished by (self-) analyzing what they have done in the situation and devising a plan to determine which elements of the situation are important and which can be safely ignored. However, Dreyfus believed the learners must become emotionally invested in what they do if they are to progress further.

Hall-Ellis and Grealy (2013) wrote, “The competent stage is essentially what cognitive psychologists refer to as problem solving.” They explained that as “rules and procedures do not simply move to the unconscious level; there is a discontinuity between the competent level and the proficient and expert levels” (p. 589). Moe (2004) described the first three stages as “detached rule-following with a high degree of deliberation” (p. 218). For the initial stages (Novice, Advanced Beginner, Competent), the decision-making process is a rational one.

However, the decision-making process shifts towards intuitiveness for the last two stages of Proficient and Expert. While the Proficient performer still needs to think about the steps to solve a problem after evaluating the situation, an Expert does not even "think," but knows intuitively what to do – given one's vast repertoires of experience gathered.

DIGITAL GAMES VS. SERIOUS GAMES

The use of games for teaching and instruction is not new: the ancient Chinese have been using *Go* – an ancient board game played with black and white pieces to train military-strategy thinking for thousands of years. Recent advancement in computing technology merely transformed how *digital* games are delivered or accessed, which is via a computing device. There are many genres of digital games according to media researchers. They include Role Playing Games (RPG), Arcade Games, First-Person Shooters (FPS), Massively Multiplayer Online Games (MMOG), and many others.

Another way of looking at games is by their purposes. For example, entertainment games are designed for fun and enjoyment. *Serious games*, on the other hand, are "games that are created primarily *not* for entertainment but for *serious* purposes" (Michael & Chen, 2005) – such as training, healthcare, policy change, and social good (De Gloria et al., 2014; Sawyer & Rejeski, 2002). Even though serious games should still contain elements of fun and enjoyment to better engage the users, their primary purpose is to optimize the dissemination of *messages*. The messages here can be either for broadcasting (as in advertisement or propaganda) or educative, pointing to their "serious" intents in improving training, healthcare practice, policy and social change, and so on.

From a training perspective, I prefer to define serious games as *virtual environments designed to provide training to a group of people, organizations, or industries*. Story-based and conversation-driven games with role-playing features are particularly popular options for digital game-based learning (DGBL) and serious games. DGBL tends to mean game projects for K-16 classroom learning and curriculum supports (All et al., 2014; Papastergiou, 2009), whereas serious games typically involve projects outside of schools. Serious games tend to tackle complex procedural tasks, such as emergency preparation, military strategy, financial skills, as well as soft skills training, such as teamwork, collaboration, decision-making, and communication.

The Serious Games Market Report (Sonawane, 2017) can provide an alternative perspective to the product. For instance, simulation training is the largest market segment for serious games followed by advertising and marketing, research and planning, human resources, and education. Each market can be divided further based on industries, including healthcare, aerospace and defense, government, education, retail, and so on. In terms of revenue, the global market for digital games was estimated at around $150 billion in 2019. In comparison, the global market for serious games was only $3 billion in 2017 and is forecasted to grow to $9 billion by 2023. Compared to the 50-year old digital game industry, the 15-year old serious games industry is but a tiny offshoot.

Cost vs. Benefit of Serious Games

So why have we not seen any serious games on the store shelves? One major problem is that game development is a very costly investment. Small-scale serious games can cost thousands, or tens of thousands of dollars, while large-scale games, such as the likes of those pioneered by the U.S. military, can cost millions of dollars! As a result, not many learning organizations can afford the means or resources to create their own (serious) games for training, learning, or testing. Many researchers and educators, therefore, either resort to using commercial off-the-shelf games *as is* (adoption) or alter the content via game *modding*[1] (adaption) for their own use. For instance, games like *Civilizations* and *Age of Empire* have been adopted for social studies and history classes, while *Neverwinter Nights* and *Minecraft* mods have been adapted for the teaching of science, programming, research, and so on.

Despite their usefulness and lower cost, game modding has its limitations. For instance, the medieval-fantasy setting of *Neverwinter Nights* makes it near impossible to create scenarios with a modernistic or futuristic setting. This is why the serious games industry came into being because they can design serious games specifically to meet the needs of target groups. In this chapter, I will limit my discussion to serious games for training, since there are other kinds of serious games, including advertisement games, games-for-health, and games-for-social-good that have very little to do with skill acquisition training.

Regular Training vs. Training with Serious Games

Several obstacles remain in the way before serious games can be more widely adopted for skill acquisition training. First, the high costs of game development means that organizations must exercise caution when investing in serious games. The cautionary investment directly affects the amount of content and game-based learning activities that can be included in any serious game. Smaller investments lead to less content, which also means the game can be completed quickly, and therefore, not enough training to yield the large effects (hoped for) in performance outcomes. Unless the stalemate can be broken, it may take a while before serious games can realistically support training as originally intended. There is hope that technology advancement and cheaper licensing models from game engine companies can positively affect the situation.

[1] Modding (from the word, *mod*-ify) – altering game content for one's purposes using the game development kit provided by the game company.

Secondly, even though serious games are becoming increasingly acceptable for training, few organizations could openly sanction their workers to play games during working hours. For instance, despite being a champion of serious game development, the U.S. Army only allows their military personnel to access training games during free time, but not regular office hours. There is still a long way before serious games are regarded as serious work due to the persistent social stigma.

Thirdly, because training with serious games can only be accessed via computing devices, not everyone can access the technology equally. For example, Byun and Loh (2015) recommended that an optimal serious game training session be around 1-2 hours. While this poses little problem to avid gamers who spend many hours playing games, non-gamers who seldom play games may require pre-training. Moreover, a small percentage of people who are not used to playing games have been documented to experience mild to severe physiological discomfort, including eyestrain/tiredness (due to prolonged staring at a computer screen), nausea (caused by fast movements in games), and even seizures (triggered by flickering lights in games and computer screen). New serious games that make use of 3D display, virtual reality (VR), and VR-goggles can have even higher incidence of nausea and seizure (Tychsen & Thio, 2020), which impedes adoption and affects public acceptance. In short, serious games may not be for everyone until some of these problems can be overcome, thus making them more suited for supplementary training activities than completely replacing training.

Over-Simplified Representation in Game Design

Life is full of nitty-gritty details, like cleaning oneself, eating meals, and using the bathroom. Game designers must oversimplify the play environments by focusing only on the essentials, while trivializing the non-essential aspects of life to make games interesting and engaging. For instance, classic FPS games focus only on taking down enemy units with a variety of weapons. The game moves a player from scenario to scenario (or level to level), showing different battlegrounds with multiple enemy units to be eliminated. The player character in the game will never feel tired or hungry because the essential feature of an FPS is to shoot at targets. There is no need to walk physically from place to place – feeling tired, hungry, or the weight of heavy weapons on one's back. Non-essential physical functions are trivialized (or removed) to focus players on the most engaging part of the game.

Serious games are no different, albeit the essential features here involve disseminating the (training) message in an engaging way to the users. A virtual hospital simulation game may focus on the identification of symptoms of patients in an Emergency Ward, while ignoring all other functions, scenarios, and wards that can be found in a real hospital. A serious game designed for diabetic children may only present scenarios to reinforce the strict timing of insulin intake, while ignoring all other aspects of a child's life. A military training game to prepare soldiers for imminent deployment in Afghanistan may depict multiple scenarios that force the trainees to speak Farsi aloud in order for the voice recognition engine to detect correct pronunciation, while another shooting simulation focuses only on target practice and not language acquisition.

Serious Games vs. Gamification and Micro-Learning

People who are new to serious games often confuse them with *gamification*. Gamification (Deterding et al., 2011) involves the application of game-design elements (e.g., leaderboard and badges) in non-game contexts to motivate user participation such as e-learning or marketing. It is completely different from serious games. In e-learning, for example, gamification can take the form of a story with a selection of choices and may *appear to be like a* *game*. Gamification in e-learning is usually accessed via web browsers, whereas serious games are "enclosed" software systems – created by game engines (e.g., Unreal or Unity3D) and are designed to provide learning/training content and activities for a specific industry or group.

Serious games may be regarded as "boutique games" commissioned by organizations to allow its trainees to practice certain skills, like learning a foreign language, training first responders for disaster preparation, recruiting young people into an industry, and so on. While it is possible that smaller game companies may attempt to market 'micro-learning' through mobile learning (games), the serious games industry typically goes after larger-scale, standalone projects that cost upwards of hundreds of thousands of dollars. Short (15-minute long) micro-learning that is trending in e-learning or gamification is not comparable in scale to serious games. Recall that an optimal training duration for serious games should last around 1-2 hours per session (Byun & Loh, 2015). This means that engaging in a mere 15-minutes of serious games is not sufficient to learn any skill meaningfully.

Instructional Technology and Skill Acquisition

The advancement of digital technologies and training methods can help to *expedite the process* of expertise development (Hoffman et al, 2013). When designed and used appropriately, serious games (as a kind of instructional technology) can benefit skill acquisition by making the training/learning materials more engaging, effective, and efficient (e^3) to the trainees/learners (Merrill, 2009).

To the uninitiated, *instructional technology* is never about digital gadgetry (as in computing technology). The domain of instructional technology is about devising *technologies* (defined as practical techniques or procedures) to facilitate, enhance, and support training/learning processes, whether or not they involve the use of media (Gagne, 1987). Let us say that the learning/training problem is about how instructors can deliver learning to a group of learners who are located in different locations. One of several solutions from an instructional technology perspective would be to prepare learning materials ahead of time and put them online for remote access – i.e., *asynchronous e-learning.* Alternately, web conference software may be used in conjunction with simultaneous video-casting (e.g., Zoom) to simulate face-to-face classroom meetings – i.e., *synchronous e-learning*. If the problem is about how to improve the quality of training/learning materials, an instructional technologist may suggest creating more e^3 (effective, efficient, engaging) learning (Merrill, 2009), or include better evaluation and feedback to re-design and improve existing processes. Serious games (as an instructional technology) can be used as the solution to address several training/learning problems; they can increase motivation and engagement, simulate dangerous environments and uncommon events for training/learning, facilitate the interactions missing in standalone training/learning sessions, collect user-generated data *in situ* gaming environment for analytics, and so on.

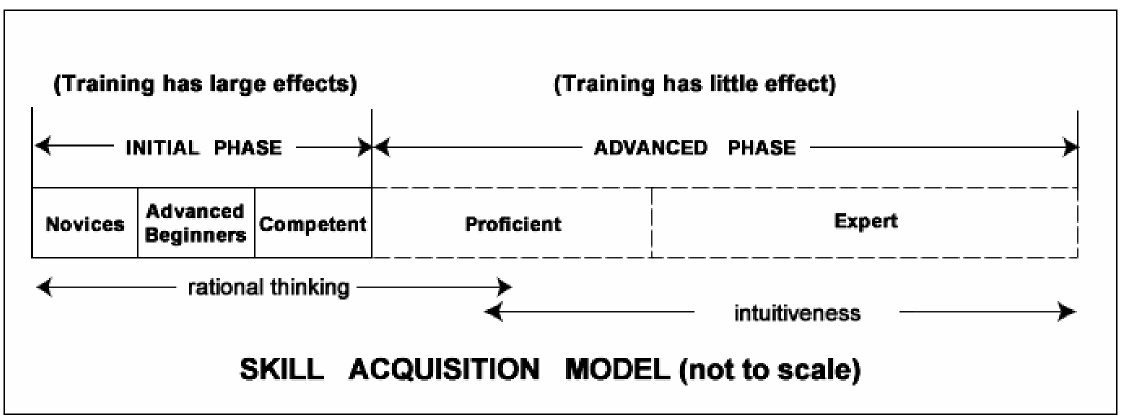
Instructional technology is seldom about designing content or curriculum for training/learning, but is usually about the design and development of tools (e.g., learning templates, mini-games, simulations, forms) to facilitate, enhance, and support the delivery of instruction. For instance, to provide better structure for online lessons, instructional technology can inform educators about the appropriate font-types, font-size, color scheme, and amount of white space to be used when designing presentation slides. In the case of serious games, instructional technology research can advise designers and educators on the *procedures* (or *technologies*) to better integrate digital games into curriculum, design effective serious games for learning, collect user-generated data via gameplay, and convert collected game data into analytics.

Trainable Phase in the Skill Acquisition Model

The Dreyfus brothers' phenomenological approach to skill acquisition (Eriksen, 2010) has successfully influenced the research about expert and skillful behaviors in many domains. In particular, the first three stages of the Dreyfus Skill Acquisition Model (2004) are most applicable in serious games (training) research from an instructional technologist's point of view. For the purpose of applying the Dreyfus Model to serious gaming, the five stages of skill acquisition can be divided into two phases. The first three stages constitute an *Initial Phase* – comprising Novice, Advanced Beginner, and Competent, and the last two stages make up an *Advanced Phase* – consisting of Proficient and Expert. Specifically, training with serious games allows learners to practice the rules acquired repeatedly (during gameplay), to faciliate the formation of long-term memeory. The first three stages are a unique phase because they have not yet crossed into the realm of emotion.

Figure 1

Skill Acquisition Model (Not to Scale): Initial (Trainable) vs. Advanced Phase



Here, the *Initial Phase* is of particular interest to instructional technologists because training can have a large effect in performance improvement – when (and if) the instructional materials are designed and used appropriately. In comparison, training has little effect during the

Advanced Phase because those in the Proficient and Expert stages already possess much experience. As leaners gather more experience, they gradually grow from rational thinking to more intuitive (Proficient), to completely intuitive (Expert) in their decision-making process. Those in the Advanced Phase improve performance through experience, or experiential training, as compared to training by instructional materials in the Initial Phase (Figure 1). In other words, serious games (as designed instructional materials) would have the greatest effect for trainees and learners who are Novices, Advanced Beginners, or Competent (of the Initial Phase), but less so for the Proficient performers and Experts (of the Advanced Phase).

PERFORMANCE ASSESSMENT WITH SERIOUS GAMES

Besides supplementary training during the Initial Phase of skill acquisition, serious games may also be employed to assess users' performance for improvement. Without assessment, there is no way to know if the learners have achieved what they set out to learn (i.e., meeting learning goals). Since the early days of serious games, researchers have asserted that without assessment, there would be no difference between serious games and entertainment games (Loh et al., 2007; Michael & Chen, 2005). Unfortunately, development of advanced assessment with serious games has been very slow.

The first obstacle is probably the cost involved. Adding assessments to serious games will further increase the cost of production because more research will be needed to ensure the assessment method is appropriate for the game. Most game companies do not have the means to support a team for game development and another for assessment (research). The former requires programmers, and the latter, statisticians or data scientists. This should not be taken to mean that performance assessment with serious games is not possible, only that it is very much in its infancy and is still limited to independent research at the moment. The second obstacle is finding the right assessment tool or method to overcome the technology barrier (in this case, the virtual environment). Unlike traditional classroom instruction where instructors can easily monitor whether students have learned the content presented, many traditional assessments become useless inside a virtual environment without physical presence. We will next examine some of the empirical research methods reported in serious games research thus far. As the rest of the chapter is about the collection of empirical data for serious games analytics, self-reported (qualitative) data will not be included.

SERIOUS GAMES RESEARCH METHODS

Pretest/Posttest Comparison and the Pitfalls of Media Comparison

Among the various empirical research approaches, pretest/posttest comparison has been shown to be the most prevalent method found in serious games research (Bellotti et al, 2013). This approach is quite commonly used in educational research and student projects (e.g., theses and dissertations) to evaluate the effectiveness of a new teaching intervention against a control group. In these studies, a pretest is first used to measure the performance baseline of participants (ensuring homogeneity) before intervention and is followed by the experimental treatment involving either the intervention (e.g., playing a game) or control. Finally, a posttest is used to measure any performance difference after treatment.

While this seemed to be a reasonable research design, Clark (1985, 1994) and colleagues (Salomon & Clark, 1974) have long criticized this approach to be flawed when used to *compare media* (and technology). Comparing media is like comparing apples against oranges and tends to yield findings that are *confounded*, which means "results are susceptible to multiple interpretations" (Cook, 2005, p. 542). Unlike medical research where the method originated, the media being compared in these experimental studies were not "identical" (or comparable) and often involved pitting traditional classroom instruction taught by teachers against another technology, such as games (Clark, 2007) or e-learning (Hastings & Tracey, 2005).

Compatible Comparison for Serious Games

A better alternative approach for serious games is the *A/B Testing method* from the field of Usability (King et al., 2017). This approach calls for *compatible comparison* using two versions (A and B) of the (same) object of study – such as creating two versions of the game from one game engine with a similar look and feel for compatible comparison, but with one (test) variation that might impact users' behaviors. For example, two versions of the same game can be used to compare the effects of guided learning (Game A) against discovery learning (Game B) (Zhou & Loh, 2020). Since the two games used in this case are of compatible comparison, they can be compared safely without fear of media comparison.

Repeated Measure Design and the Importance of Training Rounds

The Repeated Measure design method is another good alternative for serious games research. In this method, researchers would ask players to go through several rounds of the same game, but with slightly different learning goals each time. For instance, the first round of gameplay could serve as a tutorial to familiarize players of the game mechanics, keystrokes, mouse movements, and navigation, while the next few rounds of gameplay can act as iterative performance improvement. For example, players may be asked to clear a game level in progressively faster manner – e.g., first at 60 seconds, followed by 30 seconds, and finally 20 seconds. The Repeated Measure design carries an added benefit in serious games research as the participants themselves act as (an internal) control; there is no need for a control group and, hence, no worry about media comparison.

Russell-Rose and Tate (2013) differentiated *domain expertise* from *technical expertise* when it comes to performing tasks using technology. According to them, "*domain expertise* defines one's familiarity with a given subject matter," whereas "*technical expertise* indicates one's proficiency at using computers, the Internet, search engines, and the like" (p. 4). In learning with serious games, the presence of the technology barrier (game) forces learners to first become familiar with the gameplay mechanics, navigation, and game environment before they can gain access to the learning content presented. Without attaining a certain competency in using the game interface, learners will have difficulty acquiring the learning content because the technology acts as a gatekeeper to the domain knowledge. Players, first, need to achieve technical expertise through one or more rounds of tutorial/training before they can open the door to the space for skill development towards domain expertise.

This makes sense from the perspective of the Dreyfus five-stage Skill Acquisition Model (2004), as well. We know that novices require rules to learn, and the tutorial rounds provide them with the first rules of operation necessary to begin understanding how the training works inside the game environment.

Implication for Serious Game Designers and Researchers

Serious games researchers would do well to carefully consider the implication for technical expertise in designing research. This is because the first few rounds of gameplay may not reflect the learning of domain knowledge, which the researchers are interested in, because players are still trying to become familiar with (navigating) the game. It is recommended that serious games research first allow participants to achieve technical competency using tutorials before presenting the content to the learners. In fact, this is exactly how entertainment games are designed: tutorials before actual game.

When present, the tutorial/training rounds were discovered not only to significantly reduce individual differences, but also to "decrease the effects of unfamiliarity with game controls" (Young, 2014). Byun and Loh (2015) observed, "Once gender differences were adjusted for training – for example, subjecting all participants to a period of training in order for them to become familiarized with the game controls, navigation, and environment, etc., before data collection, gender differences would dissipate" (p. 135). In serious games assessment, the tutorial rounds get participants past the Novice stage quickly by removing the handicap imposed by unfamiliar technology.

SERIOUS GAMES ANALYTICS

Why would anyone need Serious Games Analytics when serious games are not (yet) ready to replace regular training? In normal research development, performance assessment usually take place after training is established (i.e., fully adopted by a community). However, due to the advent of Data Science, analytics quickly became very popular in many fields of research, including serious games. Furthermore, since the soft skills targeted by serious games training are not new, participants who already possessed the skills are readily available for analytics research. Serious games analytics researchers (like me) only need to devise a means to collect data from the players and assess their performance in serious games.

Harvesting User-Generated Data* In Situ *Serious Games

All games that take place within a game world have their own geographical (x, y, z) coordinates. As players enter that world, their locations (i.e., coordinates) are immediately monitored by the game to facilitate navigation of the players. The games not only keep track of all movement (coordinates) of the players, but also every action (time and events) that occurs. Story and role-playing games that are filled with opportunities for "player choice and action"

(Dansky, 2007) are particularly useful for this type of data collection. Players' decisions, which lead to actions performed (e.g., using an item, talking to characters, pushing a lever, entering a code, etc.), can be traced and recorded as a game log or into an online database using *telemetry* (Zoeller, 2010 or *Information Trails* (Loh et al., 2007; Loh, 2012). These data constitute a *course of actions* – an audit trail comprising the order and list of the actions performed in the game by the players, which serve as the raw materials for (serious) games analytics. Analytics can be used in data visualization to reveal what players actually do in the game during training (like a replay), thereby removing the need for video recordings and self-reported data (i.e., surveys and interviews) in serious games research.

Compared to (entertainment) game analytics (Seif El-Nasr et al., 2013) that harvests user-generated data to understand how people play for the purpose of product development and revenue generation (i.e., monetization), serious games analytics made use of the same data for user and product (performance) improvement. [A comprehensive treatise with many examples and study cases is already available elsewhere. Readers who are interested in learning more are referred to: *Serious Games Analytics: Methodologies for Performance Measurement, Assessment, and Improvement* (Loh, Sheng, & Ifenthaler, 2015).]

Gaming Skills and Levels of Expertise

Based on the descriptions from the Dreyfus five-stage Skill Acquisition Model (2004), players at different stages of the model behave differently in how they solve problems. Using a "capture the flag" military game as example, Novices tend to just charge ahead towards the goal while firing at every enemy along the way in the hopes of reaching the final goal alive. Advanced Beginners may pick off weaker enemies, but avoid stronger ones, while making sure they rest at every campground to regain health. The Competent gamers may survey the battleground first before deciding on the best strategy to capture the flag. Depending on the map, they may elect to either take out the enemies, or avoid enemy engagement altogether if the mission does not specify it. Proficient players may choose play-as-you-go and react semi-intuitively to the waves of enemies as they encounter them. When the enemies prove too tough, Proficient gamers may revert to reconnaissance to scout out the enemy formation so as to plan out their attack strategy. Expert gamers simply do whatever they do (by intuition) to capture the flag.

Depending on how a level is set up, Proficient gamers can sometimes behave like Experts if the difficulty is below their skill level; conversely, Expert gamers may also be forced to retreat and plan out their next move if the enemies proved overwhelming. In short, players' problem-solving behaviors are not only recognizable in gameplay, but also characteristics of the skill acquisition stages to which they belong. Players' gameplay *course of actions* can be either screen-recorded for *replay* towards qualitative review and analysis or traced using telemetry or Information Trails (Loh, 2012) for quantitative analysis and data mining.

Beta-Testing and Model Answers

In the entertainment gaming industry, game companies would sometimes invite gamers to "beta-test" unreleased games so as to identify and fix gameplay problems before release. Gamers from all skill levels can apply and be accepted into beta-testing programs, but for different purposes. Although gamers from any level can still provide valuable feedback to game companies on issues such as "gameplay balancing" (e.g., no over-powered enemies or weapons), level bottlenecks (e.g., no chokepoint on the map where enemies could congregate and overrun players), loopholes (no circumvention from the intended design to complete mission), and bugs (no programming errors that could crash or break the game), the real value of top-level and pro-gamers (of Proficient and Expert stages) lie in the way they play. Top-level gamers and Pro-gamers have different strategies and game-handling capabilities when compared to the general gamers, and game companies need these pro-level gameplay data to ensure *sufficient challenges* are present in these games to satisfy gamers of these categories. Similar to sport-players and musicians, members of the professional gamers league can spend 8-10 hours a day playing games to hone or maintain their skill for tournaments.

Because Proficient and Expert gamers play games rather differently from the general gamers – especially if strategizing and decision-making are involved, their courses of action (what they do) and route of navigation (where they go) can serve as the *model answers* for the game. As Expert gamers may be difficult to find (and expensive to engage), Proficient gamers may be invited to create a model answer for a game level. When secrecy is needed (e.g., to avoid media coverage), game designers themselves can substitute as Proficient gamers because they would already know the best route and actions required to quickly beat the game (i.e., they already know what the model answer looks like). In serious games for training, model answers to the game (or game level) can be obtained likewise either by asking the game (level) designers or inviting Proficient/Expert gamers to play the game several times to generate user-data and obtain the best result. Once there is a *model answer* for the serious game, its *course of actions* can then be used as baseline for comparison to rank group(s) of players based on how (dis)similar their courses of action are to the model answer. All we need then is a research tool that allows the determination of (dis)similarities between two courses of action.

String Similarity: A Primer

Early researchers from the field of Record Linkage pioneered a group of statistical methods called *string similarity metrics* to determine if two name-records are similar enough to be considered as duplicates (Bilenko & Mooney, 2003). Their aim was to clean out large databases of name-record to remove extraneous (or duplicate) data. Let us consider the two last names: *Thomson* and *Tomson*. Without a standardized metric, it would be extremely difficult to manually determine if *Thomson* is the result of a typographic error (i.e., it is the same person as *Tomson*), or a legitimate entry (i.e., a different person altogether) – hence, the need for *string similarity metric* – a method to determine (dis)similarity between two (text-) strings.

String similarity metrics or coefficients calculation typically yield a number that ranges from 0 to 1, where 0 means the two strings are completely dis-similar, and 1 means the two strings are completely similar (identical). The number may also be interpreted as percentages (i.e., 0 to 100%), if necessary. New uses in string similarity metrics have resulted in their being incorporated into applications such as fraud and plagiarism detection, facial recognition, traffic patterns analysis, and genetic sequencing.

Expert Similarity Index (ESI)

Loh and Sheng (2013b) pioneered a new string similarity metrics called *Expert Similarity Index* (ESI) to measure the (dis)similarity of the courses of actions between many players and *single expert*. They evaluated a number of string similarity metrics and found the *Jaccard coefficient* to be sufficient for general-purpose analytics (Loh, Sheng, & Li, 2015). The *Longest Common Substring* method is recommended if a more robust approach is required (Loh, Li, & Sheng, 2016). [Readers who are interested in the method are referred to the series of articles, which cover the following: (1) use of *n*-grams for “directionality of navigation” in ESI as serious games analytics: Loh & Sheng (2013a; 2013b); (2) *Maximimum Similarity Index* (MSI) – an ESI for multiple Experts: Loh, & Sheng (2014); (3) Predictive Analytics: Loh, Sheng, & Li, (2015), and (4) Training Prescription: Loh, Li, & Sheng, (2016).] I will next illustrate the ESI method with a simple example using Jaccard coefficient.

Case Study 1: Player Ranking (Identifying Best Candidates)

A few years ago, a database firm used a game to screen candidates publicly for invitations to a job interview. The backstory for the game was that a rogue robot had taken over the company, locking out everyone. The engineers of the company could create a gap in the security and slip one person into the office. Anyone who successfully found the passcode into the CEO's office and used the kill-switch to shut down the robot would be invited for a job interview at the company. Since this was before the advent of (serious) game analytics, my guess was that the database firm simply called everyone who had entered the correct password for an interview. But as a serious game researcher who is working on analytics, I felt compelled to take up the challenge to devise a solution to the problem, which led to the creation of the ESI.

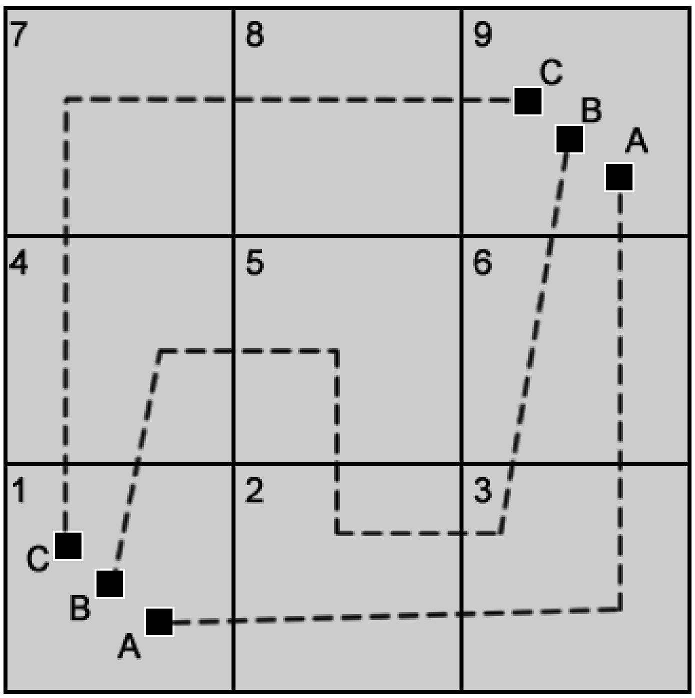
First, I created a serious game with a "Search-and-Rescue" mission and put in place the online database and game server systems needed for *Information Trails* (Loh et al., 2007). My team and I conducted the data collection and traced the course of actions of all who participated in the project. It was during my first analysis when I came to the realization that all the traced user-generated data would be meaningless unless I could compare these data with a baseline of performance. By including the *model answer* as the ideal route for baseline comparison, I was able not only to measure the performance of the players, but also rank them accordingly to identify the best candidates. More importantly, my analytics approach was scalable and could easily be applied in the performance measurement or the ranking of any number of players – from 50 to 50 000, or more. I will next illustrate this approach with an example.

Calculating the Expert Similarity Index (ESI)

Figure 2 shows a very simple 3 × 3 game map with three players (A, B, and C). The dash-lines depict the movement of the players within the game world. For example, Player A traversed the game world from square 1 → 2 → 3 → 6 → 9, while Player B took the route: 1 →

Figure 2

Course of Actions Performed by Three Players, A, B, and C



4 → 5 → 2 → 3 → 6 → 9. Player C chose the route: 1 → 4 → 7 → 8 → 9. We then convert each player's *course of actions* in the game into strings. Using Player A as example, his/her course of actions [1 → 2 → 3 → 6 → 9] can be converted into the string [12369].

Suppose we now add the *model answer* obtained from the level designer (or a Proficient/Expert gamer) to the mix where Player [E]'s course of actions is [12369]. We can then compare every player's course of actions against the model answer – the baseline of comparison, using *string similarity metrics*. Following the procedures outlined in Loh, Li, and Sheng (2016), everybody's course of action must first be converted to strings and then into *bi*-grams to preserve the directionality of the course of action.

Player [A] = [12369] → *bi*-gram [12, 23, 36, 69]
Player [B] = [1452369] → *bi*-gram [14, 45, 52, 23, 36, 69]
Player [C] = [14789] → *bi*-gram [14, 47, 78, 89]
Expert [E] = [12369] → *bi*-gram [12, 23, 36, 69]

The *string similarity index* approach calls for the comparison of two sets of strings (or bi-grams), one from the player and one from the baseline of comparison (i.e., the model answer). There are a number of metrics available for string similarity comparison, for example: the Jaccard coefficient and Longest Common Substring coefficients (formula shown below):

$$Jaccard\,[\mathrm{A, B}] = \frac{|A \cap B|}{|A \cup B|} \qquad \text{and} \qquad LCS\,[\mathrm{A, B}] = 1 - \frac{d_{LCS}(A,B)}{d_{max}\,(A,B)}$$

Here, we calculate the Jaccard coefficient between player B and the ideal route E as an example:

$|B \cap E|$ = |23, 36, 69| = 3,
$|B \cup E|$= |12, 14, 45, 52, 23, 36, 69| = 7,
Jaccard [B, E] = $\frac{3}{7}$ = .43 → ESI between Player B and ideal route.

The Expert Similarity Indices (ESIs) between the players and the model answer are:

$\mathrm{ESI}_{Jaccard}$ [A, E] = 1
$\mathrm{ESI}_{Jaccard}$ [B, E] = .43
$\mathrm{ESI}_{Jaccard}$ [C, E] = 0

We can deduce from the ESI values that Player A's performance is identical (100% similar) to the model answer. Player B's ESI is .43 (indicating that Player B's performance is 43% of the model answer), and Player C is completely dis-similar to the model answer (ESI = 0, or 0% similar).

[**Note:** Readers who are familiar with *R* (*R* Core Team, 2014) may use the *stringsim* function – found in the *stringdist* package by van der Loo (2014), to automate the calculation of string similarity. Base on my experience, there is no "one-size-fits-all" index for Serious Games Analytics. In order to identify the *best* index for a game, several indices must be compared and the best fit determined using a Jackknife cross-validation. Examples of the approach are available in Loh & Sheng, 2013b and Loh, Li, & Sheng 2016.]

Ranking of Players using Expert Similarity Index (ESI)

Once the ESIs for every player (be it 50, or 50 000, or more) have been calculated, they can then be ranked accordingly. For example, the three players (A, B, C) from the example in the previous section can be easily ranked using ESI to reveal how closely they resemble the model answer: *Player A* (1) > *Player B* (.43) > *Player C* (0).

Figure 3

Serious games analytics: Ranking of players by Expert Similarity Index (ESI)

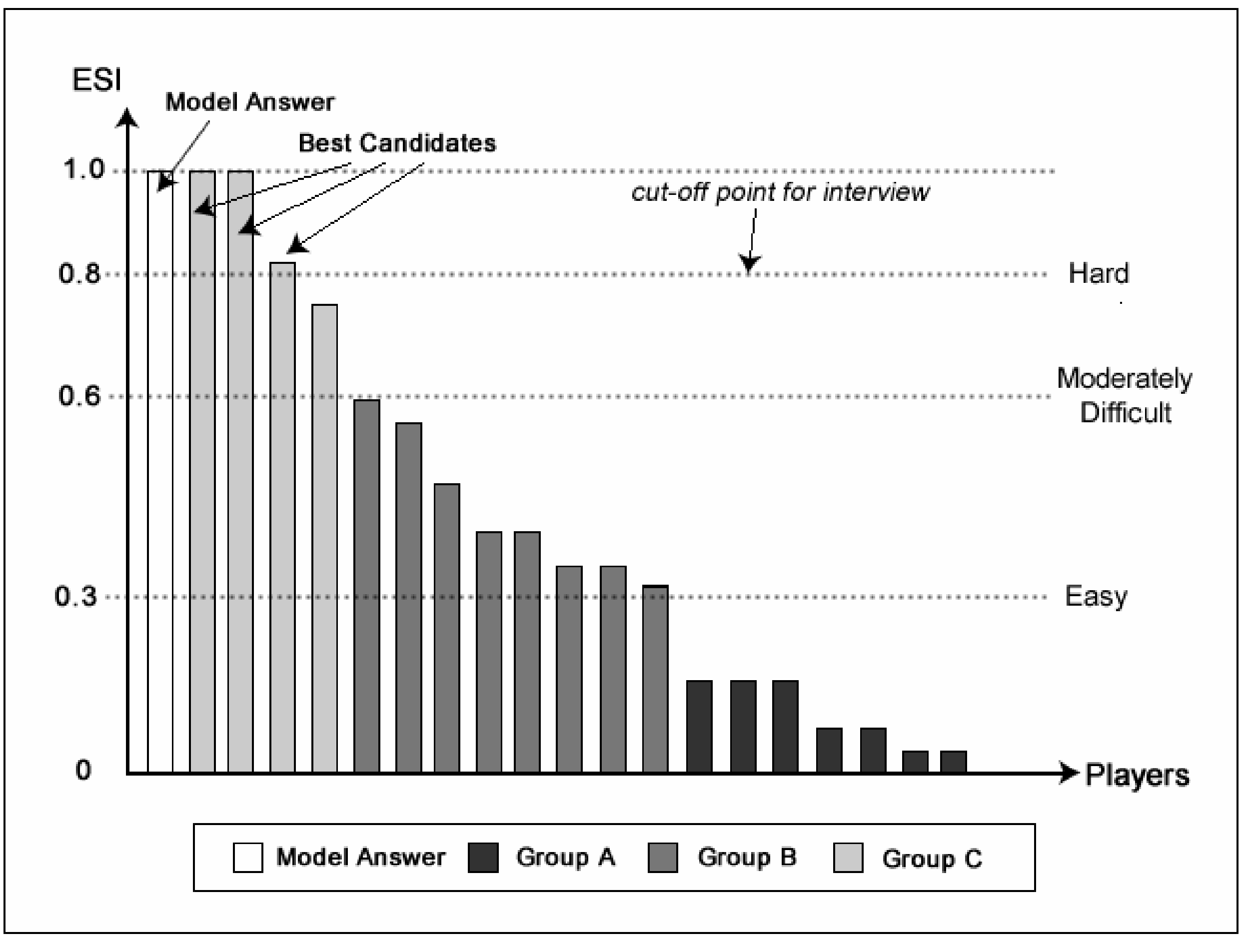


Figure 3 shows the ranking of 19 players in our project compared against the model answer (bar No. 1). A stakeholder of a company may arbitrarily decide an ESI value (e.g., .8) to be the cut-off point for invitation to interview. Figure 3 showed three arbitrarily determined cut-off points for easy (.3), moderately difficult (.6), and hard (.8). The graph showed two top-candidates (who scored an ESI of 1.0), a third candidate who scored just above the cutting-off point (above .8), and a fourth candidate who scored just below the cut-off point. It would be easy to invite the best three candidates (with one reserve) for a job interview using the analytics.

Using analytics alone, we can identify three distinct groups of players in Figure 3: Group A (high score, ESI > .6), Group B (intermediate score, .6 > ESI > .3), and Group C (low score, ESI < .3). Next, we would examine Player Behavior Profiling by taking into consideration the behaviors of the players along with their performance measurement to evaluate each group's level of skill acquisition based on Dreyfus model (2004).

Case Study 2: Player Behavior Profiling

In this case (Loh & Li, 2016), the researchers designed a serious game to observe how players solve problems. The goal of the serious game was to locate the exit portal of the maze and escape as quickly as possible. The game comprised of a maze with a starting point and two escape routes to reach a single escape portal (exit point). There was a shorter route with a locked door and a second, longer route, with no obstacle. The locked door could be opened only by solving a simple musical puzzle – by stepping on a pressure plate in front of the door to generate a musical tone to complete a musical scale: namely, *do-re-mi-fa-so-la-ti*. Players needed to step on the pressure plate one more time (total: eight times) before the door could be unlocked. The last and eighth step would not generate any musical tone but would unlock the door. Players were allowed to play the game as many times as they wanted and were also allowed to quit playing at any time: either to restart the game because they had gotten lost or to conclude the research after they met the goal. All who completed the maze received a small incentive.

Maximum Expert Similarity (MSI)

Because the serious game in Case Study 2 contained multiple model answers, the ESI method described earlier was not sufficient in meeting our needs. (ESI was made to handle only one model answer.) In serious games assessment with multiple model answers, the players' routes must be compared to *all available model answers* for similarity measures. For example, if there is a game with three model answers (1, 2, & 3), a player may score an ESI value of .3 and .5 when compared against model answer 1 and 2, but attain an ESI value of 1.0 when compared against model answer 3. In this case, the maximum score of 1.0 would be taken as the players' *Maximum Similarity Index* (or *MSI*) because, in that particular scenario, the players behaved just like the (expert) model answer. Maximum Similarity Index (MSI) is the special ESI when multiple model answers are present (see Loh & Li, 2016, for a complete treatise on this issue).

Players' Game Behavior Profiles

Participants of the project (Case Study 2) were divided into four different groups based on their in-game behaviors:

1) *Quitters* – players in this group quit playing after trying a maximum of five rounds. They showed little effort in trying to improve their performance of escaping the maze as quickly as possible. They only met the bare minimum requirement and exhibited many characteristics of non-gamers, such as getting lost in the maze and feeling confused or frustrated.

2) *Fulfillers A (Least Resistance)* – this group of players was able to escape the maze using the long (but easier) route. A few of them came across the locked door, but were unable to solve the musical puzzle. As a result, they gave up trying the puzzle, searched elsewhere, and found the second, longer route of escape. Regardless of finding the locked door or not, all players in this group settled on the long, but easier, route as the *answer*. They practiced running multiple times using this route until they could escape the maze as fast as possible (best time) before quitting. While they fulfilled the game goal, this group took more time to escape the maze than *Fulfiller B* group. However, they did not know any better because they were unaware that there was a shorter route. On the whole, this group of players spent a moderate amount of time in learning to escape the maze, longer than group 3, but shorter than group 4.
3) *Fulfillers B (More Persistence)* – this group of players was able to escape the maze with the shorter (but difficult) route. They found the short route with the door and kept working on the musical puzzle until they figured out the way to unlock the door. They then practiced running multiple times using the shortest route until they could escape the maze as quickly as possible (best time) before quitting. Like the previous group, they, too, fulfilled the game goal. They had no idea that there was an alternate, albeit longer route; they were lucky to have found the shorter route and stuck with it until the end. On the whole, this group took less time to escape the maze than the *Fulfiller A* group because they used a shorter route.
4) *Explorer* – this group of players was the most interesting because they *evaluated all available options* before deciding on the best way to escape the maze. Like some of the other players, the *Explorers* came to realize that the maze actually contained two possible

routes of escape. Unlike Group 2, they did not abandon the musical puzzle but kept working on it until they figured out how to unlock the door – in this regard, they were similar to Group 3. After finding one route of escape, they did not stop there, but continued to explore the maze and discovered a second route. The order of discovery was not important, as some found the long route first before discovering the shorter route with the locked door, and vice versa. They practiced running the maze using *both routes* (collect data) to evaluate which would be a better choice to escape the maze as quickly as possible. They only quit after they were satisfied that they had given their best performance in solving the problem. This group spent a considerable amount of time in exploration to discover all they could about the maze, but they performed just as well as the *Fulfiller B* (Group 3). They were *the only group* who learned how many escape routes there were in the game. In short, they *more than* fulfilled the game goal.

Connecting Players' Behaviors and Skill Acquisition Model

I will now compare the players' behaviors for each group against the five stages of skill acquisition as depicted in Dreyfus model. Table 1 shows a comparative summary between the players' behaviors for the corresponding stages in Dreyfus Skill Acquisition Model and groups profiled using serious games analytics.

Judging from their in-game behaviors, it was evident that the *Quitter* group was made up of new or non-gamers who put in minimum effort in escaping the maze (probably for the incentives). Since the goal (i.e., rule or instruction) of the serious game was to "escape the maze as quickly as possible," they strictly *followed the rule* – a behavior of Novices – with minimum effort. They practiced escaping the maze with little improvement and quit in less than five attempts. From another perspective, the first few rounds of gameplay acted as a tutorial for all who were new to the game, placing them at the Novice stage.

Table 1

Dreyfus Five-Stage Skill Acquisition Model for Serious Games Assessment and Analytics

Dreyfus Five Stages	**Serious Gamers' In-Game Behaviors**
	Initial Phase
Novice Follows rules learned in training. No regard for context.	Gamers strictly follow rules and instructions without further consideration of context.
Advanced Beginner Begins to recognize context of situations. Generates maxims to address situations.	Gamers show some awareness of their environment and are able to (self-) evaluate the actions and behaviors that can affect their performance. Their thinking remains simplistic – seldom beyond the immediate context of the situation. They are limited by their exposure and experience with serious games.
Competent Develops own plan for addressing what is important in situations. Becomes emotionally involved in the task.	Gamers show planning and in-depth (self-) evaluation abilities before making rational decisions. They take into consideration factors and experience beyond immediate context (i.e., have a bird's-eye view of the situation) and are emotionally involved in the outcomes of their performance (e.g., feel good if they perform well).
	Advanced Phase
Proficient Replaces rules with situational intuition, but still deliberates when making decisions. Become emotionally invested in the task.	Gamers show a mixture of planning and intuitive responses when making decisions. They are emotionally invested in the outcomes of their performance and are interested in learning how other top performers solved the problem.
Expert Reacts flexibly with intuitive, practiced understanding based on extensive experience.	These are top performers who solve problems with little planning or (self-) evaluation, but do so by intuition. They have vast experience in (serious) games and are able to handle whatever situations they are in with aplomb.

Both the *Fulfiller* (*A* and *B*) groups proceeded beyond the Novice stage with more than five rounds of practice. They went beyond simply following instructions to gaining more experience in the layout of the maze with each new attempt. Depending on their *luck*, they practiced escaping the maze using the discovered route to gradually improve their performance.

All the players finally quit after they were satisfied that they have achieved their best time in escaping the maze as quickly as possible. (Players in the *Fulfiller B* group were not much better than those from the *Fulfiller A* group in terms of their skills; they were lucky to have found the shorter route first and able to solve the musical puzzle.) Their in-game behavior placed them at the Advanced Beginner stage.

While a few of them did come across the locked door (by chance), they considered it a dead end after they were stumped and went on in search of *the* escape route. Finding the longer route further *convinced* these players that the locked door was a red herring – an erroneous conclusion due to limited exposure to games. The players' in-game behaviors and inability to think beyond the immediate context and consider the possibility of an alternate route showed them to be gamers with some skills and experiences befitting the Advanced Beginner stage.

In contrast to the preceding groups, the *Explorers* were able to think beyond the immediate context of the problem and (self-) evaluate the situation for feedback. After they discovered that there was more than one route to escape the maze, they not only took the time to run through each route, but also (self-) monitored the time taken to decide which route would afford them with a better performance. They eventually settled on the better route after extensive testing and comparison. Their in-game behaviors indicated that these players exhibited characteristics of the Competent stage, *at the very least*.

By this, I mean the game in Case Study 2 was too simple to further evaluate the players in the *Explorers* group since additional scenarios containing obstacles (such as timed puzzles and trap-evasions) requiring intuitive responses would be needed to distinguish players at the Proficient or Expert stages from the Competent performers. Since the researchers had no control over who would participate in the project, it was possible that some of the participants were

Proficient or Expert gamers. When using serious games assessment to evaluate participants' proficiencies in skills (other than game-playing), games containing specific skill-based scenarios at various levels of difficulty would have to be created; this is a possible endeavor if sufficient budget and game development resources are available. Unlike serious games for training that only address the first three stages (Initial Phase) of Dreyfus Skill Acquisition Model, it is possible for serious games assessment to be used to assess/evaluate the performance of Proficient and Expert (Advanced Phase) given appropriate design and sufficient resources.

CONCLUSION

At this point in time, serious games can be used to (a) train new skills or (b) assess performance of learned skills. The former requires much more time and resources because developers need to support the entire curriculum or program to train someone from Novice to Competent. Professional training programs can span multiple years (e.g., nursing and teacher training are four-year programs). The scale and scope of such training *using serious games alone* is not (yet) feasible/possible in today's market and technology, whereas, serious games for assessment only need to create selective *cases* to address the (five) stages under evaluation and assessment, which is much more plausible compared to training using serious games.

The adaptation of Dreyfus five-stage Skill Acquisition Model (2004) for serious games assessment and analytics is very useful because it can provide detail as to how performers of each stage behave. Such courses of action could be traced and analyzed using the Serious Games Analytics approach to determine players' skill level or stage. Serious games assessment can serve as an alternate mode of professional assessment to training communities that are already familiar with the use of serious games and Dreyfus Model, such as military, nursing, aviation, and medical training. The *Expert Similarity Indices* (ESI) and *Maximum Similarity Index* (MSI) are viable indices for measuring and empirically ranking the performance of a group of players by comparing their courses of action to a single model answer, or multiple model answers, respectively. Further research in this area may continue to expand the use of the Dreyfus Skill Acquisition Model (2004) into new domains of training and assessment from virtual games to simulations and virtual reality (VR) training.

Acknowledgement

The author would like to thank Ms. Ariel Loh for copyediting the final draft of this chapter.

REFERENCES

All, A., Nunez Castellar, E. P., & Van Looy, J. (2014). Measuring effectiveness in digital game-based learning: A methodological review. *International Journal of Serious Games*, *1*(2). https://doi.org/10.17083/ijsg.v1i2.18

Bellotti, F., Kapralos, B., Lee, K., Moreno-Ger, P., & Berta, R. (2013). Assessment in and of serious games: An overview. *Advances in Human-Computer Interaction, 2013,* 11. https://doi.org/10.1155/2013/136864

Bilenko, M., & Mooney, R. J. (2003). Adaptive duplicate detection using learnable string similarity measures. In *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2003)* (pp. 39–48). Washington DC. https://doi.org/10.1145/956750.956759

Byun, J., & Loh, C. S. (2015). Audial engagement: Effects of game sound on learner engagement in digital game-based learning environments. *Computers in Human Behavior*, *46*, 129–138. https://doi.org/10.1016/j.chb.2014.12.052

Clark, R. E. (1985). Evidence for confounding in computer-based instruction studies: Analyzing the meta-analyses. *Educational Technology Research and Development*, *33*(4), 249–262.

Clark, R. E. (1994). Media will never influence learning. *Educational Technology Research and Development*, *42*(2), 21–29. https://doi.org/10.1007/BF02299088

Clark, R. E. (2007). Learning from serious games? Arguments, evidence, and research suggestions. *Educational Technology*, *47*, 56–59. http://www.jstor.org/stable/44429512

Cook, D. A. (2005). The research we still are not doing: An agenda for the study of computer-based learning. *Academic Medicine*, *80*(6), 541–548. https://doi.org/10.1097/00001888-200506000-00005

Dansky, R. (2007). Introduction to game narrative. In C. Bateman (Ed.), *Game Writing: Narrative Skills for Videogames* (1st ed., pp. 1–23). Charles River Media.

De Gloria, A., Bellotti, F., & Berta, R. (2014). Serious games for education and training. *International Journal of Serious Games 1*(1). https://doi.org/10.17083/ijsg.v1i1.11

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining "gamification". In *Proceedings of the 15th International Academic MindTrek Conference on Envisioning Future Media Environments.* Tampere, Finland. https://doi.org/10.1145/2181037.2181040

Dreyfus, S. E. (2004). The five-stage model of adult skill acquisition. *Bulletin of Science, Technology and Society*, *24*(3), 177–181. https://doi.org/10.1177/0270467604264992

Ericsson, K. A., Prietula, M. J., & Cokely, E. T. (2007). The making of an expert. *Harvard Business Review, July-August*, 1–8. https://hbr.org/2007/07/the-making-of-an-expert

Ericsson, K Anders, & Charness, N. (1994). Expert performance: Its structure and acquisition. *American Psychologist*, *49*(8), 725–747.

Eriksen, J. W. (2010). Should soldiers think before they shoot? *Journal of Military Ethics*, *9*(3), 195–218. https://doi.org/10.1080/15027570.2010.510861

Fadde, P. J. (2009). Instructional design for advanced learners: Training recognition skills to hasten expertise. *Educational Technology Research and Development*, *57*(3), 359–376. https://doi.org/10.1007/s11423-007-9046-5

Gagne, R. M. (Ed.) (1987). *Instructional Technology: Foundations*. Routledge. https://doi.org/10.4324/9781315060248

Hall-Ellis, S. D., & Grealy, D. S. (2013). The Dreyfus model of skill acquisition: A career development framework for succession planning and management in academic libraries. *College and Research Libraries*, *74*(6), 587–603. https://doi.org/10.5860/crl12-349

Hastings, N. B., & Tracey, M. W. (2005). Does media affect learning: Where are we now? *TechTrends*, *49*(2), 28–30. https://doi.org/10.1007/bf02773968

Hoffman, R. R., Ward, P., Feltovich, P. J., Dibello, L., Fiore, S. M., & Andrews, D. H. (2013). *Accelerated Expertise: Training for High Proficiency in a Complex World*. Taylor &

Francis Group. https://doi.org/10.4324/9780203797327

King, R., Churchill, E. F., & Tan, C. (2017). *Designing with data: Improving the user experience with A/B testing*. O'Reilly. https://doi.org/10.1145/2908805.2913017

Knight, J. F., Carley, S., Tregunna, B., Jarvis, S., Smithies, R., de Freitas, S., Dunwell, I., & Mackway-Jones, K. (2010). Serious gaming technology in major incident triage training: A pragmatic controlled trial. *Resuscitation*, *81*(9), 1175–1179. https://doi.org/10.1016/j.resuscitation.2010.03.042

Loh, C. S. (2012). Information Trails: In-process assessment of game-based learning. In D. Ifenthaler, D. Eseryel, & X. Ge (Eds.), *Assessment in Game-based Learning: Foundations, Innovations, and Perspectives* (pp. 123–144). Springer. https://doi.org/10.1007/978-1-4614-3546-4

Loh, C. S., Anantachai, A., Byun, J. H., & Lenox, J. (2007). Assessing what players learned in serious games: *In situ* data collection, Information Trails, and quantitative analysis. In *Proceedings of the Computer Games: AI, Animation, Mobile, Educational & Serious Games Conference (CGAMES 2007)* (pp. 10–19). University of Wolverhampton.

Loh, C. S., & Li, I.-H. (2016). Using players' gameplay action-decision profiles to prescribe training: Reducing training costs with serious games analytics. *Proceedings of the 3rd IEEE International Conference on Data Science and Advanced Analytics (IEEE DASS 2016): Special Sessions on Game Data Science (GDS)*.

Loh, C. S., & Sheng, Y. (2014). Maximum Similarity Index (MSI): A metric to differentiate the performance of novices vs. multiple-experts in serious games. *Computer in Human Behavior*, *39*, 322–330. https://doi.org/10.1016/j.chb.2014.07.022

Loh, C. S., & Sheng, Y. (2013a). Measuring the (dis-)similarity between expert and novice behaviors as serious games analytics. *Education and Information Technologies*, *20*(1), 5–19. https://doi.org/10.1007/s10639-013-9263-y

Loh, C. S., & Sheng, Y. (2013b). Performance metrics for serious games: Will the (real) expert please step forward? *Proceedings of CGAMES 2013 USA*, 202–206. https://doi.org/10.1109/CGames.2013.6632633

Loh, C. S., Li, I. H., & Sheng, Y. (2016). Comparison of similarity measures to differentiate players' actions and decision-making profiles in serious games analytics. *Computers in Human Behavior*, *64*, 562–574. https://doi.org/10.1016/j.chb.2016.07.024

Loh, C. S., Sheng, Y., & Ifenthaler, D. (Eds.). (2015). *Serious games analytics: Methodologies for performance measurement, assessment, and improvement*. Springer International Publishing. https://doi.org/10.1007/978-3-319-05834-4

Loh, C. S., Sheng, Y., & Li, I.-H. (2015). Predicting expert-novice performance as Serious Games Analytics with objective-oriented and navigational action sequences. *Computers in Human Behavior*, *49*, 147–155. https://doi.org/10.1016/j.chb.2015.02.053

Meister, M. L. R., & Buffalo, E. A. (2017). Memory. In P. M. Conn (Ed.), *Conn's Translational Neuroscience* (pp. 693–708). Academic Press. https://doi.org/10.1016/B978-0-12-802381-5.00050-6

Merrill, M. D. (2009). Finding e3 (effective, efficient, and engaging) instruction. *Educational Technology*, *49*(3), 15–26.

Michael, D., & Chen, S. (2005). *Proof of learning: Assessment in serious games*. Gamasutra. http://www.gamasutra.com/view/feature/2433/proof_of_learning_assessment_in_.php

Moe, V. F. (2004). How to understand skill acquisition in Sport. *Bulletin of Science, Technology & Society*, *24*(3), 213–224. https://doi.org/10.1177/0270467604264996

Papastergiou, M. (2009). Digital game-based learning in high school computer science education: Impact on educational effectiveness and student motivation. *Computers and Education*, *52*(1), 1–12. https://doi.org/10.1016/j.compedu.2008.06.004

R Core Team. (2014). *R: A language and environment for statistical computing* (3.12). R Foundation for Statistical Computing. http://www.r-project.org

Russell-Rose, T., & Tate, T. (Eds.) (2013). *Designing the Search Experience: The Information Architecture of Discovery*. Morgan Kauffman. https://doi.org/10.1016/C2011-0-07401-X

Salomon, G., & Clark, R. E. (1974). Re-examining the methodology of research on media and technology in education. *Review of Educational Research*, *47*, 99–120.

Sawyer, B., & Rejeski, D. (2002). *Serious games: Improving public policy through game-based learning and simulation*. https://www.wilsoncenter.org/sites/default/files/media/documents/publication/ACF3F.pdf

Seif El-Nasr, M., Drachen, A., & Canossa, A. (Eds.). (2013). *Game Analytics: Maximizing the Value of Player Data*. Springer London.

Sonawane, K. (2017). *Serious Game Market Report*. https://www.alliedmarketresearch.com/serious-games-market

Tychsen, L., & Thio, L. L. (2020). Concern of photosensitive seizures evoked by 3D video displays or virtual reality headsets in children: Current perspective. *Eye and Brain*, *12*, 45–48. https://doi.org/10.2147/EB.S233195

van der Loo, M. P. J. (2014). The *stringdist* package for approximate string matching. *The R Journal*, *6*(1), 111–122. https://journal.r-project.org/archive/2014-1/loo.pdf

Young, M. E. (2014). Sex differences in the inference and perception of causal relations within a video game. *Frontiers in Psychology*, *5*, 926. https://doi.org/10.3389/fpsyg.2014.00926

Zhou, T. & Loh, C. S. (2020). The effects of fully and partially in-game guidance on players' declarative and procedural knowledge with a disaster preparedness serious game. *International Journal of Gaming and Computer-Mediated Simulations. 12*(4): 23-37. https:/doi.org/10.4018/IJGCMS.2020100102

Zoeller, G. (2010). Development telemetry in video games projects. *Game Developers Conference*. San Francisco, CA. http://gdc.gulbsoft.org/talk