A Methodology for Evaluating Game Technologies in Training

Jennings Baker
Lockheed Martin STS
Orlando, FL
jennings.j.baker@lmco.com

Merrill Lay
Lockheed Martin STS
Orlando, FL
merrill.e.lay@lmco.com

Kevin Reed, Ed.D
Lockheed Martin STS
Orlando, FL
kevin.m.reed@lmco.com

ABSTRACT

Important attributes of sound educational media include how interesting the media is, how expensive is the media to produce, how difficult is it to maintain media currency, and how effective is the media in transferring knowledge and skills. When viewed against these criteria, commercial computer games appear to offer significant educational opportunities. Well-designed computer games are certainly interesting; they hold the attention of players for a number of reasons: the challenge, the competition, the visualizations and the engaging storylines. Games are also relatively inexpensive and their widespread availability makes them easy to purchase. Recent trends toward on-line distributed and multi-player games provide opportunities to collectively engage large numbers of players. The power of the customer-driven marketplace ensures that games are kept current and that improvements are constantly sought as businesses try to attract larger and larger market share. The question for the training and education community to resolve is whether computer games are strictly for entertainment purposes or if they have utility in enhancing learning. This study uses an abbreviated ISD approach that supports the rapid evaluation of games and establishes their potential value for training and education applications. The study focuses on critical tasks for Tank Crewmembers by evaluating 10 commercially available games for their potential to train those tasks to some level of proficiency. The objective of this paper is to establish a methodology for evaluating commercial games. The study paves the way for future research on the design of media that leverages the engaging aspects of commercial games, but is designed to support increased human performance.

ABOUT THE AUTHORS

Jennings Baker is currently the Lockheed Martin Simulation, Training and Support Senior Instructional Engineer for Advanced Gunnery Training Systems (AGTS). In that capacity he is responsible for the training materials designed and presented on major DOD and International AGTS, Tank Gunnery Training System and Hellenic Army General Staff contracts. Jay develops and designs training scenarios and exercises, instructional course materials, training systems utilization handbooks and systems maintenance manuals. Jay served in the United States Army and retired with 21 years military service. Jay is a Gulf War veteran with personal experience on the M60A1, M60A3, M1, M1A1, M1A2 and M1A2 SEP Tanks and the Army's new Stryker Vehicle. While in the military his assignments included Tank Commander, Platoon Sergeant, First Sergeant and NCOIC of VII Corps G3 Operations.

Merrill Lay is currently a Principal Investigator in the Independent Research & Development group at Lockheed Martin Simulation, Training and Support. Merrill has over 27 years of collective experience in Flight & Ground Simulation domain, specializing in computer image generation. He was a key contributor in the development and production of the Compu-Scene series product line and held the position of Lead Product Application Engineer. His career has included assignments supporting customers in the advancement of visualization for aircraft such as the B-2, F-22, and RAH-66. Merrill holds a bachelor's degree in Electronic Engineering Technology.

Kevin Reed is currently a Lockheed Martin Simulation, Training, and Support Senior Instructional Systems Designer (ISD). In that capacity he is responsible for developing training curricula for major DOD and International Programs, to include ISD policies, procedures and tools. Kevin's previous employment includes being a USAF Fighter Training Unit F-16 Instructor Pilot. While at Luke AFB, Kevin was the Chief of Training Development for USAF and International F-16 pilot training; responsible for developing 7 distinct syllabi of instruction and for producing all academic courseware for training over 1,000 F-16 pilots per year. Kevin holds an Ed.D degree in Educational Leadership from the University of Southern California, an MA degree in International Relations from the University of Kentucky, and a BA degree in Geography from the University of Tennessee.

A Methodology for Evaluating Game Technologies in Training

Jennings Baker
Lockheed Martin STS
Orlando, FL
jennings.j.baker@lmco.com

Merrill Lay
Lockheed Martin STS
Orlando, FL
merrill.e.lay@lmco.com

Kevin Reed, Ed.D
Lockheed Martin STS
Orlando, FL
kevin.m.reed@lmco.com

INTRODUCTION

Important attributes of sound educational media include how interesting the media is, how expensive is the media to produce, and how difficult is it to maintain media currency. For a variety of reasons, well-designed computer games can satisfy each of these criteria. Games provide a challenge, they stimulate competition, the visualizations appeal to the senses, and the engaging storylines draw the players into the virtual world. Games are also relatively inexpensive and their widespread availability makes them easy to purchase. Recent trends toward on-line distributed and multi-player games provide opportunities to collectively engage large numbers of players. The power of the customer-driven marketplace ensures that games are kept current and that improvements are constantly sought as businesses try to attract larger and larger market share.

To the training practitioner, satisfying those criteria means little if the media does not effectively transfer learning. This paper establishes a protocol for assessing the training effectiveness of games. It uses an abbreviated Instructional System Design methodology to evaluate game technologies. The methodology presented supports the rapid evaluation of games and establishes their potential value for training and education applications.

Structurally, this paper presents a methodology, then illustrates the prescribed methodology by describing an investigation designed to evaluate commercially available games and U. S. Army training requirements.

Game Technology

This report focuses on games and supporting technologies that can be hosted on general-purpose PCs, including some video games that fit this definition. There are many genres of electronic games – some of which may be applicable to military training applications:

- Educational, (e.g., "Carmen San Diego")
- Role-Playing Game, (e.g., "Dungeon Masters")
- Vehicle Simulator, (e.g., "Microsoft Flight Simulator 2002 Professional")
- First-Person Shooter, (e.g., "Doom")
- Wargaming, (e.g., "Decisive Action")

Educational games are designed for self-paced learning, being generally implemented as single-player operation. Role-playing games, vehicle simulators, and first-person shooter games are often networked, to allow for competitive and cooperative play. In single-player versions of these games, the software often provides limited synthetic behaviours for potential opponents or allies in order to provide an interactive experience. Vehicle simulator and first-person shooter games are inherently video applications. Educational and role-playing games may be implemented as video, graphical, or even textual games.

Games exploit leading edge software development – including high level multi-players networking, graphics display modelling, and artificially intelligent synthetic behaviours. Some of these technologies are available as commercial products to games developers. The main advantage of game technology products, however, is their mass-market, low-cost, field-tested nature that enables game tool developers to target a specialized niche.

Today, the term "video game" is used more generally to apply to any game that integrates a continuous dynamic graphical display. It is now feasible to render this capability on today's powerful home PCs – especially with high-end graphics card add-ons. Indeed, many PC games are now offered in almost identical versions for the arcade, home console, or PC.

LITERATURE REVIEW

To appreciate the capabilities of computer games, it is instructive to understand the structure of the games. Some games are designed as "stand-alone" platforms which function without any influences on game player behaviour. Other games incorporate a type of AI, which can influence how the game unfolds. "Game coaches," for instance, are a class of software components that are "experts" in the play of a game. These intelligent tutors monitor the progress of the game in order to offer advice to the human player. Game coaches are integral to the play of the game and thus do not operate in a separate mode. Game coaches usually maintain a model of the player's level of expertise in order to provide appropriate recommendations. An example would be the online coach called WUSOR that was developed to provide advice for players of the logic-based shoot-the-hidden-monster game called "[Hunt the] Wumpus" (Carr and Goldstein, 1977). If the game coach is an imperfect advisor, however, it is often more annoying than helpful—and thus often disabled.

The current market for electronic games, consumer video, and computer games is significant—estimated at $15 billion in 2000, with some individual computer games grossing over $400 million (Gibson, 2000). When viewed in the context of each unit selling for under $50 and most of the custom hardware products to support gaming costing no more than a few hundred dollars, the requisite level of interest in games to produce such numbers of consumer is apparent.

The affordability of games has led individuals to examine games for their usefulness in training scenarios. By way of an example, in the mid-1990s Lt. Scott Barnett of the U. S. Marine Corps modified the commercial version of the three-dimensional first-person-shooter game "Doom" to create "Marine Doom." Marine Doom was used to help train four-man fire teams in urban warfare. The Marines found this game sufficiently compelling to play the game on their own free time (U. S. Army Corps of Engineers, 2003).

A significant drawback of the "inspired individual" ad-hoc approach to transforming commercially available games for specific training applications is that games which were originally intended for individual consumers have not been developed within frameworks that have been (or are in the process of being) adopted in military training contexts. These frameworks include the Sharable Content Object Reference Model (SCORM) for learning management systems and the High Level Architecture (HLA) for distributed simulations. In some cases, it may be possible to retrofit COTS games with the hooks necessary to interface into the necessary frameworks, but this is likely to be much more costly than if the interfaces were first designed into the systems. In some cases, it may not be possible to retrofit these interfaces without significant redesign. For example, it cost $750,000 to modify Spearhead II – a $30 COTS tank game – to use the HLA (Erwin, 2000).

As game technology matures, it is increasingly being employed by the commercial training industry to train users towards certification or licenses in their respective fields. For example, Lockheed Martin offered a commercial truck driver training system to Werner Enterprises in 2001; which made Werner the first trucking company to use simulation-based training for driver certification. Using the truck simulator enabled Werner to train in dangerous or unusual conditions while limiting expensive wear-and-tear on actual truck equipment. The truck simulator features day and night views that simulate a full spectrum of weather conditions, a high-definition terrain database, malfunction and road emergency response training, automated driver-performance assessment, and intelligent traffic behaviours (Pfister, 2001).

Along with commercial training enterprises and individual entrepreneurial use, institutions of higher learning are beginning to exploit the training capabilities of computer games. At the University of Southern California's Institute for Creative Technologies (ICT), partnerships between academia and the entertainment industry are being formed. These collaborations emphasize low-cost training platforms which use game-based consoles and development techniques. ICT actively engages multimedia content as well as location-based simulation and interactive gaming to exploit training content technology. Working with screenwriters from the film industry, ICT adapts concepts of story and characters to enhance a trainee's immersion experience through participation in a synthetic environment.

ICT's Mission Rehearsal Exercise Project produces a training environment filled with interactive virtual humans that engage trainees in a confusing, stressful

environment. The virtual humans are both scripted and AI driven, with a few of the characters using an AI based emotion model that interacts according to the trainee's speech inputs (Swartout, Hill, Gratch, Johnson, Kyriakakis, LaBore, Lindheim, Marsella, Miraglia, Moore, Morie, Rickel, Thiébaux, Tuch, and Whitney, 2001).

ICT's Games Project has produced both console and PC-based games such as "Full Spectrum Command" and "Full Spectrum Warrior." "Full Spectrum Command" provides leadership training toindividuals in a command and control environment, by giving a student an operational plan that requires the trainee to organize platoons, develop plans, and coordinate the actions for approximately 120 soldiers. The training scenarios are intended to develop cognitive skills for tactical decision-making, resource management and adaptive thinking. In "Full Spectrum Warrior," players/trainees assume the role of Squad Leader—receiving objectives and directing their fire teams to execute their missions. The development of Full Spectrum Warrior paid special attention to replicating actual tactics, techniques, and procedures IAW actual U.S. Army operational doctrine.

In a marriage of higher learning and private industry, LB&B Associates, of Columbia, MD, has modified the game engine from author Tom Clancy's best-selling computer game, "Rainbow Six – Rogue Spear," to train small U.S. Army units on how to prepare for a mission, work as a team during mission execution, and conduct after-action debriefs. LB&B has completed a proof-of-concept version, which has been turned over to ICT for final development (Kennedy, 2002).

METHODOLOGY

For purposes of clarification, this paper adopts certain terms that replicate larger concepts. The term "game" is used to refer to games that are played on a PC, video games (such as Xbox or GameBoy), and other games that are played over the Internet. "Target Population" refers to that group of individuals who is intended to benefit from the training provided by the game. "Experts" refers to individuals who have domain expertise (usually as former members of the target population) on the type of training that is being evaluated. An "Investigator" is the individual who conducts the game evaluation. The term "Stakeholder" refers to an individual who will use the results of the evaluation to reach some decision about the project which generated the game evaluation.

Begin with the End in Mind

Often, the investigator is also a process stakeholder. In instances where this is not the case, the investigator and stakeholder must decide early on as to what kind of data the evaluation should produce. A simple and effective method for ensuring the evaluation produces data that the stakeholder can use, is to start the investigation by creating the presentation that will be given at the conclusion of the evaluation. After the stakeholder has indicated what kind of data is needed to make a decision, the investigator should produce a template for data presentation. In spaces where the template requires data, notional data is inserted as a placeholder until the evaluation produces the actual results. By determining the format of the data *a priori*, it guides the proper conduct of the evaluation.

Without a clear understanding of exactly how data from the evaluation will be presented, it is possible to generate data that does not satisfy stakeholder requirements. For instance, a stakeholder may desire that data from an investigation be presented in a bi-modal fashion identifying games that support the training of certain tasks and games that do not support training. If by happenstance the evaluation were set up to solicit responses tied to a five-point Likert scale, responses on the mid-range score would not indicate whether training requirements were supported or not.

To illustrate, this paper evaluated the training potential of computer games for the U.S. Department of Defense. Specifically, commercially available armored combat games were assessed to determine their suitability as training media for U. S. Army tank crewmembers. After consultation with the process stakeholder, the investigator prepared a chart designed to capture the results produced by the study. The chart had candidate games as the vertical axis, and crew positions along the horizontal axis, with scores residing in the intersections of the axes. By either shading or not shading the data on the chart, a bi-modal demarcation of tasks supported and tasks non-supported was provided.

Use Appropriate Requirements

To produce a consequential study, is imperative that correct training requirements are used. To illustrate, the target population for this study is U. S. Army tank crewmembers. Accordingly, the tasks that tank crewmembers have for employing a tank in combat are the requirements used to assess the suitability of candidate games.

The preeminent source of requirements knowledge resides in a Training Needs Analysis (TNA) of that particular job function. A TNA contains detailed information about all task requirements, to include the Conditions the task is to be completed under, (e.g., Day or Night), and the Standards to which the task must be accomplished, (e.g., within 100 meters). In the absence of resident expert knowledge, a TNA can be used by itself to guide the selection of requirements to evaluate, but the optimum situation is to have an expert use the TNA to guide the selection of requirements. In the absence of a TNA, a suitable substitute is a job description, such as a U.S. Army Field Manual, or an equivalent civilian job description used in industry.

The game assessment is enhanced if the TNA contains a *Difficulty/Importance/Frequency* (DIF) *Analysis*, and a *Cue Analysis*. A DIF analysis indicates how often a task must be trained to reach a desired level of performance. For example, a task that is difficult, important, and used infrequently will require more training than tasks without these attributes. A cue analysis provides the initiating signals that prompt the user to take action. Cue analyses show if visual, aural, tactile, or other sensory inputs are required to start, persist, or terminate an activity.

Because a TNA was not available for the evaluation conducted for this paper, Field Manual (FM) 3-20-12 (Tank Gunnery Training) was used as the primary reference for determining the tasks to be evaluated (http://www.globalsecurity.org/military/library/policy/army/fm/). FM 3-20-12 lists all the tasks that a tank crewmember is expected to be able to perform in a combat environment.

The expert for this evaluation was a highly experienced veteran of armored combat. Using FM 3-20-12 as a guide, the expert identified 10 tasks for each of the three tank crew positions: tank commander, gunner, and driver. These 10 tasks represented the critical tasks required for each duty position to fight a tank in combat. Tasks identified for each duty position are listed below:

Tank Commander

- Vehicle and aircraft identification and classification
- Engage targets using the commander's caliber .50 machinegun
- Engage targets using the commander's Gunner's Primary Sight Extension (M1A1, M1A2 and M1A2 SEP) and CITV (M1A2 and M1A2 SEP only)
- Target acquisition
- Communicate using visual signaling techniques while mounted
- React to chemical or biological hazards
- Navigate while mounted
- Issue initial and subsequent fire commands
- Select fighting positions
- Prepare and submit Tactical Reports

Gunner

- Vehicle and aircraft identification and classification
- Location and function of gunner's controls and indicators
- Identify targets using the Thermal Imaging System and Daylight Channel of the GPS
- Target Acquisition. Detect a target and give a crew acquisition report
- Scanning techniques
- Engagement techniques
- Misfire procedures and Immediate Action
- Boresight procedures
- Prep-to-Fire checks
- Enter computer and environmental data

Driver

- Drive an M1 series tank
- Navigate from point to point using the steer-to indicator
- Detect a target or potential threat (Improvised Explosive Device) and give a crew acquisition report
- Communicate using visual signaling techniques while mounted
- React to indirect fire
- Seek covered and concealed routes
- Seek defilade fighting positions

- Drive while using the Driver's Thermal Viewer
- Berm Drills
- Missile drills

Assemble Experts

Expert knowledge of task requirements is the foundation for any study investigating the suitability of games as training media. Before the expert panel is convened, prepare a detailed in-brief which includes the materials that panel members will use to conduct the actual evaluation. Ideally, a practice session should be conducted using one of the candidate games, but on a task that is not to be evaluated. A guided practice session standardizes the responses of the evaluators. To standardize the administration of the evaluation, the experts should fill out the same response sheets they will use during the actual evaluation.
As the evaluation conducted for this paper was a methodology validation exercise, only a single expert was employed. The expert for this study was given an extensive in-brief, and practice evaluations were conducted in order to elucidate the scoring of responses. Practice sessions used the same scoring sheets that were used in the subsequent validation evaluation.

Establish Game Criteria

Early on, appropriate criteria for choosing candidate games has to be established. First, based on the requirements of the target population, choose games that seem likely to have some benefit as training media. If you are interested flight training, for instance, pick games that simulate piloting an aircraft. A search of the relevant literature on the games being considered will narrow the list of possible games. Due to the frequency with which games are updated in order to compete in the commercial market place, articles over one-year old should be used only in the absence of more recent articles. Knowledge of which games incorporate open-source technology can refine the search for candidate games; in that open-source games are more easily enhanced—and thus more appealing.

As the evaluation conducted for this paper was a study of tank crewmember tasks, games that used the premise of armored combat were considered. Once an initial list of games was developed, gaming web sites were reviewed for more insight on how the games were actually played; to include analyzing any hints-for-advanced-play, which might provide insight into which games to evaluate. Specialists with expertise in the open-source gaming domain further distilled the list of potential armored combat games to evaluate.

Based on the reviews described above, 10 games were selected for evaluation:

- Spearhead
- Operation Flashpoint
- M1 Tank Platoon II
- BattleField 1942 Desert Combat
- M1A2
- Panzer Commander
- America's Army
- Armored FIST 2
- Armored FIST 3
- Steel Beasts

Evaluate the Games and Collect Data

During the conduct of the evaluation, the investigator should be available to answer any questions. Quite often, questions which arise during the conduct of the evaluation were not considered when the evaluation was set up. If unanticipated questions arise, the investigator should interrupt the evaluation to ensure that all evaluators are made aware of any new information.

Most often, evaluators will be filling out a form that was prepared especially for that particular evaluation. Typically the form will contain blocks for the evaluators to record their appraisals. If written comments from evaluators are desired, then the format for these comments should have been stressed during the in-brief. Before the evaluators are released, it is important that the investigator read the comments and make sure that any ambiguities are resolved.

Analyze Data

If evaluations were recorded in numeric format or bi-modal fashion, a simple table makes it easy to compare scores. The best case scenario for the evaluation is for one game to demonstrate the highest potential for training all the individual tasks. Typically, though, one game will score highest for one training task, while another game is deemed to show more potential for training another training

task. In this instance, multiple games could be used to train multiple tasks, or the game that scores the highest in the most categories is the game that should be developed for training multiple tasks.

For "Not Supported" Determine Enhancements

More than likely, some of the games will not support the training of some of the tasks. In these instances, the game may still prove to be capable of providing effective training if the game is enhanced by changes to the game engine or software that configures the game. Games with open architecture are most easily modified. Usually, the only way to determine if it is feasible to enhance the game is to open up the game engine and read the software code. This will require access to software programmers—ideally programmers who have experience with computer games.

If it appears to be cost effective, use the enhancements to improve the game. After the game has been modified, it is necessary to reevaluate the game to assess the training effectiveness of the improved game.

Validate Results with Target Population

Once experts have evaluated the games, the results of the analysis need to be validated by having a sample of the target population perform a similar evaluation. The optimum validation consists of both numeric scores and verbal comments. Comments are especially important when the results differ widely between the panel of experts and the target population. Comments can be taken back to the panel of experts to help resolve the discrepancies. If the discrepancies cannot be resolved, the opinion of the target population takes precedence. Scores from the target population need to be recorded using the same scale that evaluators used. If warranted, the results can be placed into a bi-modal distribution of games that support the training of some tasks and games that do not support the training of some tasks.

RESULTS

10 games were evaluated for this study. Each of the 10 games was assessed for their potential to provide effective training for 3 crew positions: tank commander, driver, and gunner. Each of these 3 crew positions was defined by 10 task requirements. Each of the 10 requirements was graded on a 0 (non-effective) to 10 (highly effective) scale. Accordingly, a totally non-effective game equates to a score of 0, while a 100% effective game equates to a score of 100 (10 tasks, each scored as 10).

Based on the expert's assessments, games were placed in a bi-modal distribution of tasks supported and tasks non-supported through use of shaded blocks in the chart. There was no preordained forced-distribution of game results—the evaluation could assess all the games, or none of the games, as having the potential for providing effective training. Table 1 captures the scores for each of the 10 games.

Table 1. Tasks Supported and Tasks Non-Supported

Game	Tank Commander	Gunner	Driver
Operation Flashpoint	83	18	59
M1 Tank Platoon II	65	56	0
Armored FIST 2	32	32	45
Armored FIST 3	35	32	45
Spearhead	27	11	8
M1A2	12	9	14
Panzer Commander	9	3	9
BattleField 1942 Desert Combat	20	0	0
Steel Beast	0	8	0

Based on the evaluation, two games demonstrated the highest potential for training tasks to some level of proficiency. "Operation Flashpoint" was judged to provide the best training for the Tank Commander

and Driver crew positions, while "M1 Tank Platoon II" provided the best training for the Gunner crew position.

Although "Operation Flashpoint" does not have all the attributes necessary to effectively train armored operations, (i.e., it does not train the Gunner crew position well) it showed the most capability of the games evaluated. It was also determined that "Operation Flashpoint" could be easily modified to include vehicles with realistic instrumentation in order to improve the baseline game

SUMMARY

This paper presented a methodology for evaluating the suitability of commercially available computer games as training media. A repeatable process based on abbreviated ISD principles has been described and demonstrated that takes a practitioner from game selection through game validation. Properly used, the process described in this paper results in informed decision-making.

REFERENCES

Carr, B., & Goldstein, I. (1977). *Overlays: A theory of modeling for computer aided instruction*. Retrieved November 15, 2003, from Massachusetts Institute of Technology, Artificial Intelligence Laboratory Web site: http://cms.mit.edu/games/education/proto.html.

Gibson, Nick. (2000, January). Games Investor Online. "The online games market." Retrieved December 10, 2003, from http://www.gamesinvestor.com/Thinkpieces/The_Online_Games_market/the_online_games_market.html.

Erwin, S. (2000). Video games gaining clout as military training tools. *National Defense Magazine*, [Electronic version]. November.

Pfister, N. (2001). "Truckers to train like jet fighters." *Orlando Business Journal*, May 3.

U.S. Army Corps of Engineers. (1996, August). Marine Doom. In *Commercial Terrain Visualization Software.* Retrieved November 15, 2003 from http://www.tec.army.mil/TD/tvd/survey/Marine_Doom.html

Swartout, W., Hill, R.W. Jr., Gratch, J., Johnson, L.W., Kyriakakis, C., LaBore, C., Lindheim, R., Marsella, S., Miraglia, D., Moore, B., Morie, J.F., Rickel, J., Thiébaux, M., Tuch, L., & Whitney, R. (2001). Toward the holodeck: Integrating graphics, sound, character and story. *Proceedings of 5th International Conference on Autonomous Agents, Montreal, Canada.*