Validation of the Virtual Leopard 2 crew commander virtual reality training device

Proposed demonstration of an assessment approach

Brad Cain
DRDC - Toronto Research Centre

Defence Research and Development Canada

Scientific Report
DRDC Toronto TM 2013-030
October 2014

Validation of the Virtual Leopard 2 crew commander virtual reality training device

Proposed demonstration of an assessment approach

Brad Cain
DRDC - Toronto Research Centre

Defence Research and Development Canada

Scientific Report
DRDC Toronto TM 2013-030
October 2014

Abstract

This document outlines the proposed strategy to empirically validate the training effectiveness of a computer game-based virtual environment, Virtual Leopard 2, developed by the Canadian Armed Forces Army Learning Support Centre. The objective is to assess the training effectiveness quantitatively, comparing its use during several Crew Commander qualification courses, and make an estimate of the cost effectiveness of the use of simulation in this training application. A review of the literature indicates that empirical validation of the training effectiveness is not routinely performed despite feasible methodologies. A secondary objective of the proposed training effectiveness study is to promote the use of validation techniques for novel instructional methods to a wider audience within the Canadian Armed Forces training community. Validation exercises of this type increase confidence in the level of cost effectiveness of training provided and may be used to create a business case that has empirical support. Such studies also increase awareness of how technologies can support training as well as where they are not as effective and alternative solutions may be more appropriate.

Significance to defence and security

The proposed study will provide the CAF with empirical evidence that can be used both to assess the utility of game-based software training methods as well as to justify future development plans. The study will provide a template for evaluating other, similar studies that will contribute to the scientific body of knowledge that in turn can be used to make informed decisions on the appropriateness of using selected technologies or methods for various training purposes. Studies such as this allow DND/CAF to justify development and acquisition of emerging training technologies using factual, cost-benefit evidence along with subjective opinion.

Résumé

Le présent document décrit la stratégie proposée pour valider de manière empirique l'efficacité en matière d'instruction d'un environnement virtuel basé sur un jeu informatisé (Virtual Leopard 2) créé par le Centre de soutien à l'apprentissage de l'Armée de terre (CSAAT) des Forces armées canadiennes (FAC). L'objectif est d'évaluer quantitativement l'efficacité de l'instruction offerte par cet environnement en comparant les résultats de son utilisation dans le cadre de cours de qualification pour devenir chef de char, et d'estimer la rentabilité de l'utilisation de la simulation pour cette instruction. Selon un examen de la documentation, la validation empirique de l'efficacité en matière d'instruction n'est pas effectuée régulièrement malgré des méthodes faisables. Une autre des objectifs de cette étude de l'efficacité d'instruction serait la promotion à plus grande échelle au sein de la communauté d'instruction des FAC de l'utilisation de techniques de validation pour les nouvelles méthodes d'instruction. Les exercices de validation de ce genre augmentent le niveau de confiance relativement à la rentabilité de l'instruction offerte et peuvent servir à rédiger une étude de rentabilité étayée par des données empiriques. De telles études favorisent aussi la sensibilisation sur la manière dont les technologies peuvent aider à l'instruction et démontrent les points faibles de leur utilisation en plus de présenter des solutions potentiellement plus appropriées.

Importance pour la défense et la sécurité

L'étude proposée fournira aux FAC des preuves empiriques pouvant servir à évaluer l'utilité de méthodes d'instruction logicielle basée sur des jeux ainsi qu'à justifier de futurs plans de conception. En outre, l'étude sera un modèle pour évaluer des études similaires contribuant à l'ensemble des données scientifiques qui pourront être utilisées pour prendre des décisions informées sur la pertinence de l'utilisation de certaines technologies ou méthodes à différentes fins d'instruction. Les études comme la présente permettent au MDN/aux FAC de justifier le développement et l'acquisition de nouvelles technologies d'instruction en s'appuyant sur des faits, des preuves relatives au rapport coûts-bénéfices, et une opinion subjective.

Table of contents

1 Introduction

The Canadian Armed Forces (CAF) and the Department of National Defence (DND) are modernizing training (CDA, 2011a, 2011b), examining new technologies and approaches to determine whether they are appropriate to provide more capable, effective or affordable training. In general, the CAF[1] and are looking to new technologies:

- to provide cost effective means of delivering capability
- to extend CAF capabilities to meet anticipated challenges
- to reduce the effect of operations on the environment
- to meet the expectations of CAF members by adopting evidence-based best practices

The use of simulators has gained widespread acceptance for training in many domains despite the lack of empirical evidence that they represent an effective means of training in those domains. In the early days of simulation based training, there was considerable skepticism about the validity of simulation; now the pendulum seems to have swung to the other extreme and simulations seem to be accepted without question. There is little doubt that simulation and computer based technologies such as Virtual Reality (VR) have important roles to play in training and education (Lee, Wong, & Fung, 2010), yet determining which roles for which they are both effective and efficient should not be taken for granted (Muller et al., 2009; Stedmon & Stone, 2001). Development of training systems based on these technologies, while often cheaper than their real world counterparts, remain expensive undertakings if they are to be effective (Chatham, 2007).

Chatham (2007), programme manager for the U.S. Defence Advanced Research Projects Agency (DARPA) on training superiority, offered insights into the use of game-based training that is relevant to many forms of computer-based training. Some commonly believed myths about training games, and by extension computer simulations for training, that need to be dispelled are that they are:

1. cheap to create, deploy and maintain
2. fast, providing instant development and delivery
3. effective for automatically transferring training to competencies in the real world
4. trainer-less, so users can learn unsupervised
5. universal, so anybody with a PC can use them

A great deal of work by teams with diverse areas of expertise is required to make computer simulations cost-effective for training.

[1] VCDS. (2006). DAOD 2010 Modelling and Simulation. (DAOD 2010). Government of Canada Retrieved from http://www.admfincs.forces.gc.ca/admfincs/subjects/daod/2010/0_e.asp.

The CAF Army Learning Support Centre (ALSC) is responsible to developing computer-based models and simulations that support the CAF Land Force Doctrine and Training System (LFDTS). They have developed a high-resolution, three dimensional (3D) model of the interior of the CAF Leopard 2 within the Unity3D Game Engine[2] (a modelling and simulation tool for creating three dimensional and behavioural synthetic environments.) This Virtual Leopard 2 (VL2) provides selected functionality within the simulated vehicle for each of the four crew stations (Crew Commander, CC; Gunner, GNR; Loader, LDR; and Driver, DRV.) The VL2 provides an alternative to physical simulators to exploit advances in computer technology and simulation that has the potential to increase the flexibility and cost-effectiveness of simulation based training should this VR approach prove effective.

VR can be described as a mosaic of technologies that support the creation of synthetic, highly interactive, three dimensional (3D) spatial environments that may represent either real or non-real situations. VR has been proposed for training for many years because of its technological characteristics such as: creation of 3D spatial representations; multisensory channels for user interaction; immersion of the user in the Virtual Environment (VE); intuitive interaction through natural manipulations in real time (Mikropoulos & Natsis, 2011). The number of articles purporting the benefits of these characteristics is legion (Bell & Fogler, 1995), yet all too often, these characteristics are merely assumed to provide training. Too few empirical studies have been conducted to documented enhanced post-test knowledge or skills of students using more immersive approaches over more traditional instruction or other, less immersive, computer based approaches (Dalgarno & Lee, 2010).

In many respects, however, the questions should not be whether people can learn using one instructional medium or another: people seem to learn in even the most inhospitable environments. More importantly: “Can what is learned in a VE transfer to the real world equivalent?” If the answer is YES, then a subsequent issue arises: “Is the VE an affordable method of teaching this task compared to other methods of equal effectiveness?” In other words: “Is a VE training solution cost-effective for its proposed purpose?” It is rather pointless to pursue a training solution unless it is both affordable and effective.

Unfortunately, determining cost-effectiveness will vary between applications and over time, as technology costs decrease and as knowledge of effective uses increases. A universal, broad spectrum endorsement of VE technologies for training and education seems unwise at this point, until substantially more case studies are documented to define the limits of applicability. This is not an admonishment to not explore VR technologies, simply that uses should entail empirical validation of training effectiveness.

This report is a working document to support the development of a research plan to assess the effectiveness of simulator-based and VR-based technologies for team training within the CA. The report is intended to provide background on the role of realism, fidelity, immersion, presence on learning as well as to note areas specific to the Leopard 2 crew training environment being considered in this study that would benefit from DND/CAF stakeholder clarification. Thus, there will be portions of the report that are incomplete, indicating a need for further consideration by stakeholders as the project is refined.

---

[2] http://unity3d.com/

1.1 Objective

The objective of the VL2 study is to assess alternatives to current methods of training armoured vehicle crews to determine whether training can be accomplished using VR and game technology to replace some of the existing live and simulator based training approaches. Additionally, estimates will be made of the cost effectiveness of the training methods used in the study. The study proposes to assess desktop simulation based training and VR training approaches that are thought to provide more "immersion" to see if there is any benefit to either over the current field training approach.

1.2 Approach

The proposed approach is to use Crew Commander (CC) qualification as a demonstration case to conduct a study of the training effectiveness. Basic qualification for the CC role was identified by ALSC and the Armour School to be a suitable demonstration case for training as the CAF are resuming training for new tank crews after a focus on reconnaissance since approximately 2006. The study will be limited to the training and assessment process associated with new CCs in order to provide a manageable work package. Similar methods to assess the technologies for other basic crew qualification training or more advanced tasks could also be developed.

The proposed study will use selected processes, procedures and drills relevant to qualifying for the CC role as outline in the associated Qualification Standard (DND, 2009) and Training Plan (DND, 2006). Training scenarios that mimic the traces that are currently conducted in the field during qualification of prospective CCs will be used to provide related experiences in an SE. The proposed assessment will be based on training effectiveness to the extent practicable rather than subjective opinion. While subjective opinion can be useful, particularly when provided by subject matter experts (SMEs), they are prone to unintended biases; relying on objective, quantitative assessment according to the scientific method provides a more rigorous and defensible assessment of the training effectiveness.

Both physical simulator and VR simulation conditions are proposed for the study. Interest in physical simulators of various levels of fidelity for training is growing within DND, as demonstrated by proposed capital projects such as the Land Vehicle Crew Training System (LV CTS). Including both physical and virtual simulators in a study may provide insight that will help formulate such large projects by providing empirical data upon which to base related cost-effectiveness estimates.

1.3 Simulator effectiveness for training

Clark (1994) notes that most studies of educational effectiveness of various delivery mechanisms confound the media with instructional method and fail to control for what is being taught. The use of computer generated simulation or simulators are typically developed to replace real world training for a variety of reasons, but they are still approximations to the real world that can bring both advantages as well as disadvantages for learning effectiveness and efficiency. Others (Stoffregen, Bardy, Smart, & Pagulayan, 2003) feel that more useful metrics of training effectiveness can be developed from measurements of performance, rather than the subjective experience of fidelity. Performance measures focus on the outcomes while subjective assessments focus on the impressions; the former is what is of interest to the instructors, while the latter is what typically gets the attention of the developers and providers of simulators.

Assessment of simulator effectiveness and cost effectiveness are based on the learning rate and the transfer of training from the simulator to the real world application. Dede (2009) considers 2 types of transfer: sequestered problem solving (near transfer), which is working on real world problems that are very similar and directly identifiable with the learning environment (also includes tests of knowledge); and, future learning (far transfer), which is learning to learn to apply knowledge to novel situations that are superficially quite different from the learning environment but have deeper characteristics that are semantically similar despite being distinct. Both types of transfer are relevant to military training, the first showing competency in a domain and the second indicating mastery typically gained through extensive experience.

There are a number of metrics of simulator effectiveness (AGARD, 1980; Lathan, Tracey, Sebrechts, Clawson, & Higgins, 2002; Roscoe, 1971; Roscoe & Williges, 1980) and the choice of a suitable effectiveness metric depends upon the experimental design and the data that can be obtained in the study. Care must be taken to ensure that testing is “unbiased” so that the comparison among conditions is justified. Unbiased in this context refers to the many ways that human experimentation can inadvertently skew the results.

The Percent Transfer is a common metric used to assess how effective simulator training is in learning a task and it may be a useful snapshot to assess a single, proposed solution compared to a traditional method using operational equipment where most of the cost of training rests in the use of the operational equipment. The Percent Transfer metric may be expressed as:

$$PercentTransfer = \frac{TimeTo\Pr oficiency_{ControlGroup} - TimeTo\Pr oficiency_{SimulatorGroup}}{TimeTo\Pr oficiency_{ControlGroup}} \qquad (1)$$

where *TimeToProficiency* refers to the amount of time (or attempts) required to reach an acceptable performance level in the operational equipment. In this case, the *Control Group* would usually be students following a traditional method of training in the field while the *Simulator Group* would be students who trained in a simulator prior to moving to field training.

However, Roscoe and Williges (1980) noted that the Percent Transfer metric does not adequately discriminate among various alternatives as it ignores the time that the *Simulator Group* has to invest in the simulator training in order to realize an improvement over the *Control Group.* Roscoe and Williges recommend a metric that provides detailed assessment of effectiveness; either the Cumulative Transfer Effectiveness Ratio or the Incremental Transfer Effectiveness Ratio. The Cumulative Transfer Effectiveness Ratio (TER) metric is assessed by:

$$TER = \frac{TimeToProficiency_{ControlGroup} - TimeToProficiency_{SimulatorGroup}}{Time_{Simulation}} \qquad (2)$$

where the numerator is the same as for the Percent Transfer relationship but the denominator, $Time_{Simulation}$, represents the time spent training in the simulator. The Incremental Transfer Effectiveness Ratio (ITER) is assessed by:

$$ITER = \frac{TimeToProficiency_{(t+\Delta t)SimulatorTime} - TimeToProficiency_{(t)SimulatorTime}}{\Delta t_{SimulatorTime}} \qquad (3)$$

where the numerator now is the difference in time required by a single group to reach proficiency in the real world after receiving training in the simulator but with durations differing by the amount of the denominator $\Delta t_{SimulatorTime}$.

The TER is somewhat abstract and difficult to interpret in a meaningful way and the ITER is somewhat expensive to determine as it requires multiple training tests with different groups exposed to different simulator session durations. Typically, the TER is less than 1, indicating that training in the simulator is not as efficient as in the operational equipment. This is not universally true and it is often not the case during initial learning, but the incremental benefits generally diminish with practice as proficiency is attained.

However, even when the TER is substantially less than 1, training in the simulator may be of net benefit to both the student and the organization, assuming improvement continues, if the cost of simulation time is less than the cost of operational equipment time. This leads to a proposed Cost Effectiveness Ratio (CER) that describes the relative cost of achieving a specified level of proficiency between one approach (usually a proposed “improved” training method) and another (usually a control condition such as the traditional training method). When comparing a combination of simulator and field training to traditional field-only training, the CER may be defined as:

$$CER = \frac{Cost_{FieldTraining} TimeToProficiency_{SimulatorGroup} + Cost_{SimulatorTraining} Time_{Simulation}}{Cost_{FieldTraining} TimeToProficiency_{ControlGroup}} \qquad (4)$$

where each of the *Cost* factors is expressed per unit of time spent in the associated training environment.

Either the TER or the CER may be used to assess the merits of a simulator augmented training program relative to the traditional (control) training program, depending upon the factor that is considered most important: Time to proficiency; or, Cost of proficiency.

Few VE systems have been tested for cost effectiveness. Caird (1996) notes that VEs costs were originally high, but development cost is increasingly becoming less of a factor as technology and software improve, largely driven by the entertainment market. Unfortunately, Caird also notes that many skills are difficult to acquire in VEs precisely because of many of the advantages that make VEs inexpensive – the user interface is largely computer generated – and that "…clever inclusion of cost-effective VE components that capture the essential user-task-environment relationships is the challenge that faces VE builders on a budget." The techniques that are valid for entertainment are not necessarily the techniques that are valid for training or analysis.

It is informative to have successive evaluations of performance as the students learn in order to determine typical learning rates. This will allow prediction of when further training is unlikely to provide substantial gains in performance. Successive evaluations provide feedback, both to the organization and the student, about the rate of improvement. Finally, it is often helpful to evaluate the performance of qualified personnel in the simulator environment to assess how much adaptation to the virtual environment is required to achieve the desired standard. Methods of instruction for experts may differ considerably from that of novices, but the expert should still be able to readily accommodate to the novice training if it is a valid approach.

1.4 Review of some relevant literature

The use of simulator based training is widespread within the aviation industry and aviation simulation was largely justified by the high cost and risk associated with training in actual aircraft. The success of aviation simulation for training has sparked interest in exploiting simulation for other training, but simulation has not always been a cost effective option for many applications. This has changed with the development of game-based computer technologies. Today, much of the hardware and software supporting simulation developed for the entertainment industry makes an affordable alternative to live training with operational equipment. Whether the desired learning transfer to the operational environment occurs remains to be determined in most cases, as the literature, particularly the aviation simulator training literature, contains examples of success and failure for effective learning with these technologies (Mestre & Fuchs, 2006).

Learning in virtual environments has been documented in a number of studies. For instance, Zyda (2007) conducted a study of inquiry based learning using 3 Multi-user Virtual Environments (MUVEs) and a control group with 2000 students. The results indicated that students benefited from VR training, but only when coupled with Intelligent Agent (IA) instruction. In cases of group or teacher guided instruction in the MUVE, performance was equivalent to the traditional, passive, classroom instruction control group. No effect sizes were reported, but graphical data suggest that the differences between the best and worst were small despite being significant (presumably due to the large sample size). Such examples highlight the need for well controlled studies that are subsequently assessed not only for effectiveness, but also for efficiency of instructional method. This may also be an example where the novelty of interacting with new technology (in this case, Intelligent Agents) may be responsible instead of the principal factor (in this case, the MUVE) for the observed effects.

A.1.1 Subjectivity and validation

Subjective assessment based on aspects of realism seems to be the norm for acceptance of simulations and simulators, but such criteria do not guarantee that training in the device is effective or that training occurs at all. Research has shown that realism is a poor yardstick for measuring the effectiveness of training devices and that a disciplined instructional design approach that matches technological capabilities to training needs is essential to ensure proper adoption of training capabilities. Indeed, Clark (1994) states that learning is influenced more by the content and instructional strategy in a medium than by the type of medium itself. Clark observes that the medium for instructional delivery provides surface features that will affect cost but not effectiveness; instructional methods are the necessary, structural features that provide the opportunity to learn. However, different instructional approaches may be more efficient with selected instructional media for reaching a desired level of expertise (Mikropoulos & Natsis, 2011); not all approaches are equally cost-effective.

In order to accomplish this matching of training goals and results, the outcome of a training needs assessment of the domain and the validation of the training effectiveness of the technology must be first established (Stone, 2001). The subjective opinion of Subject Matter Experts (SMEs) both in the application domain and in the training domain about the usefulness of a training device is important for identifying potential solutions; however, such opinions are subject to biases. The validity of their opinions should be established through rigorous scientific evaluation of training effectiveness to avoid trial-and-error approaches that characterized the early days of aviation simulator training.

Dalgarno and Lee (2010) suggest that there are a number of underlying assumptions related to the enthusiasm behind adopting VE/VLEs for training, many of which have not been shown to be universally true and some have not been studied extensively at all. Some of these **assumptions** are:

1. Learners will trust their VE-based experiences sufficiently to modify their mental models, while correcting any misconceptions already held.

2. Factual information that is learnt within a 3D VLE will result in greater transfer of learning to the corresponding real environment.

3. The greater the fidelity of a 3-D VLE the greater a sense of presence will be developed and consequently, greater transfer.

4. Interactivity provided by 3D VLEs will result in greater spatial learning than would occur when passively viewing and equivalent animation.

5. 3D Multi-user Virtual Environments (MUVEs) representational fidelity and the embodied actions they facilitate will result in richer online identity, constructing a greater sense of co-presence, and that this in turn will bring about more effective collaborative learning.

Unfortunately, many developers of VLE/MUVEs assume that these are proven, universal truths and fail to verify that these outcomes are realized, or even consider whether desired outcomes could be achieved more simply with another medium of delivery. Relying on instructional methods that select the most appropriate medium for delivery seems to take a back seat to fitting instruction into the latest technology.

Fortunately, the principled application of Human Factors (HF) scientific methods to obtain validity measures as outline previously provides a means to rigorously predict the effectiveness and efficiency of training devices and methods. This equates to the Kirkpatrick Levels 2 and 3 validation assessment (Kirkpatrick & Kirkpatrick, 2006) that could subsequently be incorporated in longitudinal studies to determine organizational outcomes (changes in overall effectiveness) that typify Kirkpatrick's Level 4 assessment.

A.1.2 Constructs and training

The concepts of simulation fidelity, realism, immersion and presence are frequently cited as prerequisites for effective learning that employs simulation (Baum, Riedel, Hays, & Mirabella, 1982; Dalgarno & Lee, 2010; Dede, 2009; Nichols, Haldane, & Wilson, 2000; Patel, Bailenson, Hack-Jung, Diankov, & Bajcsy, 2006). This phenomenon began in the early days of simulators that has carried on into VR and is based in part on hypotheses from Constructivism learning philosophies[3]. It should be highlighted that these concepts predate VR and do not depend upon VR, but rather are an interaction between the user and the presentation medium.

These hypotheses were largely speculative without much supporting evidence from the education science literature (Lombard & Ditton, 1997; Nichols, et al., 2000; Persky, et al., 2009). Throughout the years, voices warn against relying on intuition and advocate verifying, through scientific study, to determine where VEs were useful and where they were not useful for training (Caird, 1996; Lee, Wong & Fung, 2008). In the interim, little empirical support has emerged for a strong causal relationship between these concepts and training effectiveness or learning when they are manipulated directly (Nichols, et al., 2000); these constructs may be simply correlates that may or may not arise in effective training simulations.

Other features that arise from the technological characteristics of VR that are thought to contribute to positive learning outcomes (Mikropoulos & Natsis, 2011)are: first-person experiences (Constructivism Learning, Stanton, 2006 ; Winn, 2003), natural semantics (ecological validity), transduction (using the VE as a transducer to mediate an experience), reification (transformation of abstract ideas into perceptible representations) and autonomy (of the user in the SE). Whether researchers take into account these features for the design of Educational Virtual Environments (EVEs) and whether these features do contribute to positive learning outcomes remains to be seen in many cases. Further, engagement with the 3D VR and positive reports from either students or instructors should not be confused with effective learning demonstrated by empirical measures of performance (Salzman, Dede, & Loftin, 1999).

Evaluations of EVEs frequently consider opinion but often neglect measures of performance. Usability problems with inappropriately designed user interfaces may interfere with learning, particularly by distractions within the VE itself. Subjective assessments indicate that many users like the VE and remain engaged in the experience although without objective measures of performance, the training effectiveness is ambiguous. In fact, in agreement with Clark (1994), many studies that have compared performance in VEs with other, traditional approaches found

[3] Constructivism learning theory: http://en.wikipedia.org/wiki/Constructivism_%28learning_theory%29

little or no benefits of several unique features of VR such as haptic and motion feedback (de Winter, van Leeuwen, & Happee, 2012; Mikropoulos & Natsis, 2011).

A.1.3 Fidelity and realism

Fidelity and realism are often referred to as technology or engineering concepts of correspondence between the real and simulated environments. Fidelity and realism are used interchangeably colloquially and there is no universal agreement on their distinct definitions. It has been proposed that fidelity be considered the domain of technology application. In this sense, fidelity is an assessment of how successful a simulation is when representing the associated real world characteristics. In principle, fidelity can then be measured, but figuring out what and how to measure fidelity is not a trivial exercise. Conversely, if realism is considered to be a psychological perception of a SE, then it must be inferred from indirect measures of behaviour or performance, or by subjective, self-reports. The perception of realism drives the technological push for fidelity. While fidelity and realism may be useful concepts for specifying how to build or to characterize a synthetic environment, it is presumably the psychological interpretation of the environment that is important to learning, if indeed either is important.

In general, there is an over focus on realism with the belief that more leads to better training, perhaps a misinterpretation of the original intent of Thorndyke's "Identical elements" principle. This belief is not supported by the empirical literature; at least, not in its entirety (Baum et al., 1982; Nash, Edwards, Thompson, & Barfield, 2000; Noble, 2002; Persky et al., 2009; Smallman & Cook, 2010; Welch, 1999). Stoffregen (Stoffregen et al., 2003) cites Moroney et al.[4] who found no significant difference in performance transfer between subjects trained in an FAA approved flight simulator and those trained in a desktop PC flight trainer. Estock et al. (2009) reported that although military pilots training in an air-air simulator felt that a wider field of view display was superior and provided more similarity to the aircraft, changing the field of view resulted in no significant performance difference. These are just a few examples of the differences that can arise between subjective opinion and observed performance.

As Stone (2001) notes "...rather than confound the user's performance by trying -- and just failing -- to create a highly realistic virtual environment...present the user with only those task elements (abstracted from the results of task analyses) that have relevance to the skills one wishes to foster and transfer to the real world." This opinion is shared by Nobel (2002), particularly for novices who lack the experience to understand which cues should be attended to in order to acquire the desired skill. While it is apparent that some correspondence between the training environment and the application environment is necessary and desirable, selecting the cues and critical features for representation in the training environment depends upon both the task and the student’s level of expertise, requiring the application of principles embodied in instructional design. But while there is considerable experience with instructional design, its application to VEs is still rather limited, so study is still required to verify that the similar principles apply as with other training methods.

---

[4] Moroney, W.F., Hampton, S., Biers, D.W. & Kirton, T. (1994) The use of personal computer-based training devices in teaching instrument flying: A comparative study. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 38, No. 1, pp. 95-99). SAGE Publications.

Proponents of realism often cite the naïve position that training has to be as close to the real thing as possible to be effective and to prevent negative transfer, however, Stewart et al. (2008) reviewed the literature and concluded that true negative habit transfer is rear outside of the laboratory. Their research indicated that low cost simulators can be effective if used properly and if proficiency-based training was followed instead of pre-set exposure durations so that training effectiveness can be assessed and instruction adapted over time. Salas (Salas, Bowers, & Rhodenizer, 2009) noted that all too often, funding emphasizes the technologies but not the underlying understanding of how best to use the technologies to achieve effective training. This leads to an over-emphasis on the tools with little regard for how much complexity is really required to get effective yet affordable training.

Hochmitz and Yuviler-Gavish (2011) suggest that Thorndyke's "Identical elements" principle (the degree to which the simulation looks, sounds and feels like the actual task) may be a misinterpretation and that cognitive fidelity (an attempt to invoke similar cognitive processes by maintaining equivalent stimulus-response characteristics of a task) is more important to obtaining a high degree of training transfer in domains where situation awareness and decision making task dominate. Indeed, Nobel (2002) observes that more fidelity may even be detrimental to effective learning, particularly for novices, when features of the training environment compete for attention with the objects intended for instruction. Smallman (Smallman & Cook, 2010; Smallman, Cook, Manes, & Cohen, 2007; Smallman & St. John, 2005) as well as others, have demonstrated that increasing realism can lead to impaired performance. In retrospect, this should not be surprising as training environments are predominantly approximations of the real world equivalents that are created explicitly to promote a better learning environment. Thorndike and Woodworth's concept of the link between similar elements and training transfer[5] is often carried to excess in a mistaken belief that more similarity leads to more training transfer (or Naïve Realism as Smallman has named it.) Nevertheless, some level of association between the training environment and the operational environment must exist to support the trainee's recognition of appropriate responses to various stimuli. The question for the training system then becomes "How much (fidelity, realism, etc…) is required to achieve an affordable, effective training environment that promotes transfer of learning to the operational environment?".

Cognitive fidelity is an alternative concept that does not necessarily rely on physical similarity, rather it attempts to invoke the same cognitive processes required in the real world task by maintaining the same stimulus-response characteristics of the task irrespective of the source of the cues used to stimulate the subject (Hochmitz & Yuviler-Gavish, 2011). Such an approach is particularly important when training procedural skills that reflect knowledge of how and when to perform the procedures needed to accomplish a given task as these skills develop as a result of practice through repeated exposure to relevant stimuli. Hochmitz and Yuviler-Gavish conducted a psychomotor procedural study that found training in real world was better than simulation. However, they observed that physical fidelity training in the simulation resulted in subsequent faster task performance while cognitive fidelity training in the simulation resulted in subsequent faster error correction. They concluded that all training was better than no training, but suggest that the effectiveness of a particular simulation approach will depend upon the characteristics of the task to be learned so that a generalization of effectiveness is unwarranted.

[5] Transfer of learning: http://en.wikipedia.org/wiki/Transfer_of_learning

A.1.4 Presence and immersion

Immersion and presence are concepts that often used interchangeably, although some differences may be implied (McMahan, 2003). Both concepts are related to the observer's perception of the simulated environment and association with the corresponding real environment, although as with fidelity and realism, immersion and presence may be usefully differentiated as technological and psychological aspects of experiences in SEs. Lee (2004) defines presence as "a psychological state in which virtual (para-authentic or artificial) objects are experienced as actual objects in either sensory or nonsensory ways", although it is colloquially defined as the "sense of being there, in the virtual environment." Immersion, while also reflecting a sense of being in the VE, may, according to some, imply more than a single perceptual channel is being stimulated by the VE, whereas presence does not have this restriction.

Perhaps the most useful distinction between immersion and presence is that immersion is a characteristic of the SE, describing which senses are stimulated artificially and thus can be, in principle, measured; whereas, presence in a psychological construct that can only be inferred from self reports or performance models. If, however, we do separate immersion and presence as two distinct concepts, the first technology oriented and the second psychologically oriented, we are in a position to assess whether the technology is appropriate for task training independent of the question of whether it affects the psychological construct of presence.

In a well written paper, Welch (1999) notes there is little solid evidence to indicate that increased presence causes increased performance, although it may well be the case that the two phenomena are correlated, both being similarly affected by changes to the VE. Lee (Lee et al., 2010), however, suggested that there is a causal relationship between presence and performance based on Structural Equation Modelling evidence, although this meta analysis was based on studies that did not actually control for presence, merely measured it coincidentally. Welch (1999) conducted a study intended to manipulate presence by varying the number of sensory stimuli and cues presented to the subjects in a driving simulation. The resulting self-report of presence was greater with increased immersion (increased number of sense cueing represented); however, performance did not differ significantly amongst conditions. Winn et al. (Winn, Windschitl, Fruland, & Lee, 2002) note that although educators assume that the directness of the experience (the illusion of being there in the operational domain and the natural interaction with the environment) will lead to better learning, "there has been little systematic study of this assumption." They go on to observe that the studies that do exist provide scant evidence that immersion makes it easier to understand complex problem spaces, although the studies do show evidence of learning. This suggests that the use of immersion should be based on a cost-benefit assessment and used where appropriate, rather than being the go-to solution because immersive simulations are often more costly to develop than are many other instructional methods.

Persky et al. (2009) followed an experimental strategy similar to that of Welch using an active (able to physically move in and interact with an immersive environment) Virtual Learning Environment (VLE) to collect relevant information from a series of lectures provided by virtual agents. Performance was compared that obtained in a didactic version of the VLE, where subjects were seated and passively received the information using the same short lectures. Persky also found increased presence in the active VLE compared to the didactive VLE, but found no difference in the learning achieved, as measured by a quiz after each session. This led them to conclude that increased presence does not equate to increased learning.

Knerr (2006) concluded, based on the literature from a number of studies conducted at the U.S. Army Research Institute, that the usefulness of immersive VEs for training dismounted soldiers for spatially dependent tasks is positive but that the effects are a small increment above similar but non-immersive VE training. Knerr noted that there is a lot of enthusiasm about the utility of games for training and that they are popular with the users. Thus, while there is some evidence that immersive games can be used effectively, there is little evidence that indicates they should be used (i.e., that an immersive game is a cost effective alternative to other training approaches.).

Salzman et al. (Salzman, Dede, & Loftin, 1999) observed that many tasks are becoming more complex, requiring users to understand complicated information spaces to find patterns in the information. They hypothesize that such tasks require the use of sophisticated representations of relevant information to interpret the information and communicate understanding with team mates. They propose that immersive VR is a technology that can enhance learning of complex concepts, some beyond our normal senses. Viewing events and phenomena from multiple frames of reference, particularly dynamic simulations that involve a number of "moving parts", in a more familiar geometric representation may make some problems tractable that would otherwise be difficult to comprehend from a single perspective. It is hypothesized that, with sufficient and varied experiences in dynamic, immersive VR environments, users can develop generic mental models that will transfer to real world applications; unfortunately, it is not yet clear which features of VR lead to useful learning and which features merely distract the user from the intended learning objective (Salzman, Dede, & Loftin, 1999; Salzman, Dede, Loftin, & Chen, 1999; Smallman & Cook, 2010).

A.1.5 Three dimensional presentations of information

VE/VRs are similar to other types of training simulators except that additional user interfaces are employed and much of the user's interaction is immediate. VEs typically comprise three-dimensional, computer generated scenes to support cognitive and physical interaction (Caird, 1996) in much the same way as other immersive simulators, however, more of the 3D environment is virtual rather than physical. The distinction is somewhat arbitrary.

A few studies do exist that indicate that immersive VEs are beneficial for learning complex, three-dimensional, dynamic phenomena, a description that reflects many of the demands of a vehicle CC. Immersive VEs appear to promote more exploration of the VE because of the natural viewpoint achievable with immersive, head-tracked, head mounted displays, but this may also present perceptual challenges for changing viewpoints from ego-centric to exo-centric while immersed. Immersive VEs have also been observed impair communication and explanation between immersed and non-immersed participants, which has implications for collaborative learning and VE design, particularly for dispersed participants. VEs provide pedagogical opportunities that allow the instructor to manipulate visual perspectives, experiences *et cetera*, but whether this is more effective than other instruction methods remains a research topic for many applications.

Ketelhut et al. (Ketelhut, Nelson, Clarke, & Dede, 2010) conducted a study of students learning electromagnetic field relationships, a topic that is 3D and that had been identified as difficult to master with conventional learning approaches. They compared 3 field of view (FOV) perspectives: exo, ego and bi-centric (able to switch between exo and ego views at will), but did not include a control group in a more traditional, didactic approach. They found that students learned more when they learned using both perspectives together (bi-centric), but they did not learn differently between the exo and ego perspectives. Ketelhut also suggested that the bi-centric group seemed to have a trend to correct misconceptions somewhat easier than the other, single perspective views, supporting the view that the flexibility of using VR may be useful for enhancing learning but that the optimal use of the technology may not be as intuitive as one might believe.

A.1.6 Situated learning, constructivism and training approaches

VR has been proposed as a more affordable means of learning facts and relationships from information-rich, situated (virtual) environments following an active learning approach. The hypothesis is that students will develop an appreciation for the contextual value of the information while also developing an understanding of the underlying, abstract structure that allows generalization to novel situations. Zyda (2007) suggests that such active learning is more time consuming but that it is more effective than passive assimilation of lecture based, didactive instruction. In spite of the intuitive appeal of situated learning in context of use, this goal has proven difficult to actualize in traditional, classroom instruction because of the difficulty reproducing the contextual environment; VR addresses this problem (Barab et al., 2007).

Situated learning and constructivism are mutually supportive learning theories (two of numerous theories used to explain learning, Stoffregen et al., 2003) that emphasize the roles of the student and experience for developing understanding and expertise. Extreme positions on these two philosophies posit that learning is entirely up to the student and that any learning that is of practical value has to occur in an environment that closely matches the application domain, prompting some to note that the main weakness of situated cognition is, it seems, precisely its “situatedness” (Bereiter, 1997). Fortunately for the training community, it appears that substantial deviations from the application domain may still permit useful learning and the role of the instructor, at the very least, can affect the efficiency if not substantially influence the effectiveness of instruction. Nevertheless, elements of both of these theories are common in many instructional programs, both practical and theoretical.

Situated embodiment (where the student is situated narratively and perceptually, with a goal, an active role and engages in actions) involves more than the learner seeing a concept or even context of use. Situated embodiment involves being in the context and recognizing the value of concepts as tools useful for understanding and solving problems central to the context (Barab et al., 2007). Situated embodiment in curriculum typically involves a storyline that provides (a) legitimacy to the content and student actions, (b) a meaningful goal and set of actions for the learner, (c) a background against which learner actions have some consequence, and (d) a contextual framing that allows the learner to appreciate the use-value of the content being learned. Barab et al. (2007) conducted studies of science education with children using a VLE. In the preliminary study, they found that their SE had too much detail: students were clearly engaged with learning to play the game and superficial understanding occurred (near transfer), but there was little evidence that they developed generalizable concepts applicable in other situations (far transfer).

The importance of Barab’s work for adult education is the distinction between expert and novice. Interpretation of scientific data and generalizations must be made cautiously, and assumptions examined carefully for relevance to target applications, without making broad generalizations if cost-effective solutions are to be realized. For instance, interpreting the work of Barab et al. (2007) with children may be appropriate in general for an adult context if the basic characteristic (domain knowledge novice versus expert) holds, but effect size may vary between the two contexts because of other differences between the groups such as motivation.

Perception is selective and we cannot comprehend all of the detail that is presented to us (Stoffregen et al., 2003). A good deal of detail that is available to our senses is ignored; some is integrated subconsciously into a conceptual interpretation of reality. Only a small amount of what we can sense actually is consciously processed. A major concern relating to the use of simulation for training stems from the difficulty of determining, in many cases, the simulation's adequacy for training purposes and what information is needed to stimulate the conscious processing (which typically comprises the teaching objective) while retaining enough contextual information that missing information does not raise the automatic perceptual processes to the level of awareness. Thus, high-fidelity “situatedness” may not be necessary for creating an effective experience. In fact, as Caird (1996) observes a complex setting is likely to divert attention and thus reduce the effectiveness of the training; visual complexity may, however, be used to highlight important features to be learned.

A.1.7 Community of practice suggestions

Although there is a lack of empirical evidence, researchers in the field do have suggestions about how immersive EVEs should be used. Winn et al. (2002, p. 498) suggest considering 3 elements when designing an instructional VE to create an effective learning environment:

1. Challenge is greatest when the goal is clear, initial uncertainty is high and the activities necessary to attain the goal are of intermediate difficulty.

2. Curiosity is aroused when students believe that interacting with the game will provide the knowledge they need to have, when the game is not too complex to discourage the student nor to simple to be boring.

3. Fantasy arises when the student can imagine a number of possible outcomes to the activity and as possibilities are eliminated (through self exploration, guided instruction, feedback, etc.) only one possibility remains: the solution to the problem.

Stedmon and Stone (2001, p. 678) have suggested a checklist of considerations to help guide the development of proposed EVEs:

1. Will it improve the effectiveness with which knowledge is delivered or assimilated?
2. Will it reduce the reliance on scarce operational systems or costly hardware-based training material?
3. Does it offer anything over and above conventional training methods?
4. Can previous investments in technology be protected, or must new (custom) systems be procured?
5. Will students and trainers actually use the technology?
6. Will there be a positive transfer of training or knowledge from the computerized setting to the real operational environment?

Salas et al. (2009, p. 199) recommend focusing on “the design of human-centered training systems that support the acquisition of complex skills" rather than realism and that "in the quest for a more realistic simulation, we may have lost sight of the true goal--a more effective training device". This changes the EVE design process from one that is technology oriented to one that focuses on “those parts of the task situation which are necessary for learning to perform the task” (Muller et al., 2009) and the technologies that support those task elements.

Chatham (2009, p. 217) suggests that training is enhanced when learning and its subsequent application occur within a short interval. Affordable, personal computer based training has a greater potential to allow this to happen compared with large installations; incorporating intelligent tutoring within these systems may further emphasise their utility, augmenting one-on-one tutoring under controlled conditions that reduces the demand for good tutors, who in turn struggle to remain current with the lessons learned from the field in today's continuously changing, complex operational environment.

Dahlstrom et al. (2009) note that such complex operational environments likely imply that it is infeasible to train for all the specific situations. This, efforts should be made to create training environments that can be used to train generic competencies as well as procedural skills. In fact, emergency procedures training is largely generic in military operations as the uncertainty in evolving situations makes risk assessment and decision making an art rather than a science, much as is hypothesized in Recognition Primed Decision Making (Zsambok & Klein, 1997). Flexible, simple systems may thus be better at teaching resilience: thinking outside the box or the ability of a system to recognize, absorb and adapt to disruptions that fall outside a system's design base. Dahlstrom et al. suggest that lower fidelity simulations are cost effective solutions that may actually improve many aspects of learning that help people deal with unanticipated situations precisely because the lack of realism forces students to rely on imagination and initiative, particularly in the areas of CRM, where procedures often are inadequate when encountering novel events. Dahlstrom et al. acknowledge the usefulness of VR solutions for training, even citing examples where they are more appropriate than their real world equivalent training devices. The important aspect in the design of the VR solution, which actually applies to most training solutions, is not a question of realism as much as that the tasks to be trained must be carefully analyzed so that the simulation can be tailored to the task demands.

2 Method

2.1 Experimental strategy

The proposed study will involve two stages. The intent is to exploit regularly scheduled CC course serials to minimize disruption and costs, however, the study may require multiple serials to obtain sufficient numbers of subjects in each experimental group to have a reasonable expectation of getting statistically valid results.

The first investigation is a Reverse Transfer of Training (RTOT) assessment to evaluate the ability of the simulations to teach the intended skills and knowledge required to qualify as a Leopard 2 CC. This investigation will involve qualified CC or instructors to perform the tactics expected of the prospective CCs in the simulated environments until they meet the required level of proficiency. Ideally, these qualified personnel should first demonstrate their proficiency in the field, completing the traces the student perform and be assessed similarly according to standard. The intent of this assessment is to establish the range of variability in performance and to minimize differences in interpretation of the standards across the evaluators.

Evaluation of students in the simulated environment also takes place in the RTOT. The student CCs practice the traces in the simulation until they reach the proficiency criteria. The results of the Expert and Student group performances are compared to assess the simulations' ability to train the CC skills as a prediction for the second stage of the study. The first stage is also used to identify problems with learning in the simulations prior to the second stage of the study.

The second stage is a Forward Transfer of Training (FTOT) study using a cross-over design where the students that participated in the RTOT would then perform to criterion along with a control group who only trained in the field, following the current instructional method. This is the usual form of training validation studies, but it is difficult to accomplish in practice due to constraints and lack of experimental control associated with field studies. Taken together, RTOT and FTOT provide complementary insight into the appropriateness of the training approach intervention.

2.2 Classroom training

The study approach will attempt to minimize disruption of the existing course content beyond the specific manipulations, retaining the existing classroom and field training aspects of the current course. All subjects from each group will receive the same conventional classroom CC theory with the assumption that simulation would not be expected to replace this portion of the training programme.

2.3 Teaching points from Performance Objectives

There is a significant amount of lecture material from the CC course, sections 301-304, that is tested through a written exam. A passing grade of 60% is mandatory prior to field portion of the course.

The field training portion of the current course consists of three high level capabilities as outlined in Appendix C of the CC QS (DND, 2009):

CC a tank in offensive operations (PO.301)

CC a tank in defensive operations (PO.302)

CC a tank in transitional operations (PO.303)

Much of the focus of the CC course is on offensive operations and the associated strategies or tactics, but many of the underlying tasks to perform the strategies are common across the capabilities. Although details may differ, the receiving of orders, conducting the commander's appreciation, crew briefing and preparation for the mission follow common processes in all phases of operations. Similarly, receiving orders, passing information and requesting permissions by radio follow standardized protocols.

Skilled tasks such as detection and identification of enemy or effective use of terrain may be difficult to accomplish adequately in the SE due to limitations on visual displays. SEs do not have the computational capability to reproduce the rich set of visual cues that are available in the real world; fortunately, not all of the cues are necessary to achieve useful training. However, if some of the cues are not present, or are misrepresented, then judgements learned from experiences in the SE observing those cues may not transfer into the field or may lead to negative transfer.

Simple visual identification of enemy assets should be feasible using gross features of objects in the SE; learning to detect enemy assets under combat conditions may not be appropriate due to impoverished level of detail typical of SEs. Also, judgements of depth in SEs is known to be compressed (de Winter, Wieringa, et al., 2012; Macedonia & Rosenbloom, 2001; Peli, 1998; Pollock, Burton, Kelly, Gilbert, & Winer, 2012; Smallman, St. John, & Cohen, 2002; Watt, Akeley, Ernst, & Banks, 2005; Willemsen, Gooch, Thompson, & Creem-Regehr, 2008), so visually determining range to target would be inappropriate for simulation based training. However, training that focuses on the identification of the threats and selection of appropriate responses (indication of position, radio communications, relaying commander's intent, etc.) may be appropriate in the SE.

Similarly, relative positioning within the fire team or squadron during various Order of March (OOM) and the rationale for selecting specific OOM under certain conditions is likely to be appropriate but attempting to teach spacing may not be for computer based training (CBT) within an SE. Some conceptually simple but operationally difficult processes such as taking a suitable firing position during a bound or a cresting drill may be feasible in the SE if it relies on gross terrain features. The issues, processes and commands associated with moving from position to position (commander's intent, orders to DRV, Turret Down, Sights Up, Hull Down, assignment of arcs, etc.) should be feasible to teach even in impoverished visual scenes of typical SEs.

Although there has been advancement in automatic gesture recognition, its use is not widespread in simulation environments and implementations are custom. This suggests that incorporating hand and arm signals into a training simulation may be technologically costly and better left to alternative forms of training. In other words, proposed uses of technology should be judged and assessed a priori for their suitability for the intended application rather than attempting to accomplish everything within a single framework.

2.4 Experimental manipulations

As noted, the “as-is” control training condition is the train-to-competency in the field, although the number of repetitions is limited. It would be useful from a methodological perspective to relax the limitations on repetitions for the purpose of the study so that equivalency of the different training methods can be established.

The first manipulation to the “as-is” CC training course is practice of the CC role performing traces similar to those performed in the field, but using a VR simulation with minimal physical interfaces. To all other feasible, the training environment will be created virtually and presented to the student visually through a head mounted display. Crew members will also be simulated as semi-intelligent avatars so that crew safety and coordination tasks can be assessed. Interaction with the virtual crew members will be by voice using speech recognition and speech production software, and the CC will be able see the virtual crewmember avatars. Some tactile devices may be required to represent the CC’s sight control, maps and notes used in commanding a tank during the trace. Other vehicles in the fire team and squadron will likewise be represented virtually.

The second manipulation to the “as-is” training is a more traditional desktop simulator for the CC workstation. The same SE will be used as in the VR simulation, however, visual display of the outside world will be provide by an array of limited field of view flat panel displays. Crew members will be simulated using the same semi-intelligent agents with speech communication, but there will be no visual representation as an avatar. The tactile devices and additional equipment will be the same as for the VR manipulation.

2.5 Apparatus

A.1.8 Synthetic environment

Two distinct user interfaces to a common SE are proposed for the study: an immersive VR simulation and a desktop simulation.

The immersive VR simulation will provide minimal infrastructure:

- Visual display through a tracked HMD providing unlimited field of regard (NB. This could be VR goggles, a fully occluded HMD, or an HMD with Fused Reality capability depending on the objectives of CTC)
- manual control for the CCs gun control

For the purposes of the study, consumer computer gaming devices will be used wherever practicable. A single Virtual Reality VR2000 stereo HMD with head tracking was purchased to assess its suitability for the experiment. This HMD can deliver an XGA resolution image (1024x768) 3D display to the user.

The desktop simulation will use the same SE and controls, but the outside view will be restricted to a set of panoramic computer displays to provide approximately 120 degrees FOV or less to reflect the capabilities similar to the LCGT.

Two different solutions were purchased for the speech recognition hardware. Two Andrea noise cancelling microphone headsets (NC-181 VM USB High Fidelity Monaural) were purchased to assess their suitability as examples of commercial-gaming level microphone hardware. Additionally, two Countryman professional microphones (H6 heaset) were purchased as a similar version had proven effective in a previous study (Cain, Magee, & Kersten, 2011). Two Sure X2U XLR-USB Signal Adapters were purchased to connect the Countryman microphones to the computers for subsequent speech processing.

A.1.9 Field environment

The apparatus for the field study in the forward transfer will comprise using the operational equipment on the CFB Gagetown ranges, as per the normal training procedure. The envisioned data collection will require additional equipment to record video and sound for each vehicle in the troop.

Twelve GoPro™ Hero 3+ Black Edition cameras[6] and auxiliary equipment have been purchased by DRDC for the study. These cameras are lightweight (227g including secondary “Backpack” battery and waterproof housing), weather proof and capable of recording video for the duration of each trace. They are capable of recording 60 frames per second with a resolution of 1080p for approximately 1.5 hours on a single battery or 2.5 hours when used with the auxiliary battery. Included are helmet straps similar to that used to mount Night Vision Goggles (NVG) on the CAF issue helmet. These cameras will provide the capability to record the visual scene of the wearer for later assessment, including the perspectives of the student and the DS for each vehicle in the troop.

The tank environment is noisy, so recording of verbal communications with the GoPro cameras is unlikely; further, there is no means to capture radio communications by directly connecting the cameras to the communications system. Eight Marantz Solid State Recorders (PMD661 MkII) digital audio recorders were purchased and will be outfitted such that they can record radio and intercom communications in each vehicle. These recorders are capable of storing approximately 2 h of monaural data at a modest sampling rate (equivalent to an MP3 recording at 128 Mbps.).

2.6 Subjects

QS indicates that the CC course capacity limits are 4 and 16 with an optimum of 8.

QS indicates that all CC students will have the following prerequisite knowledge:

a. NCM DP 3 Armour Reconnaissance Crew Commander (AIWJ)

b. Officer DP 1 Armour Reconnaissance Troop Leader (ACZU)

c. Leopard C-2 Crew Commander Gunnery (AJHI)

---

[6] http://gopro.com/support/hero3-black-support

As a serial typically has more students than can be accommodated in the field at the same time, we would try to get half of the students who were left behind to do the Reverse portion of the study in the simulator during their week off. They would then be the experimental group in the forward portion of the study.

The other half of the serial could be the forward transfer control group without the simulator.

A problem is that the students only get 3 shots at reaching proficiency then they are failed. It would be better if there were an unlimited number of attempts and they practiced to proficiency (a pass) but perhaps that is unreasonable.

2.7 Experiential learning

The CC course incorporates a field portion of practical study that provides instructors with an opportunity to assess students on a basic set of CC capabilities. Not all of the practical field portion of the course will be suitable for simulation (for example, the commander's appreciation exercise); however practice receiving orders and conducting traces should be feasible within a simulation.

A sample of the current scenarios for situational briefing and an attack advance trace could be incorporated with the advice of SMEs the Armour School to create a simulation of the experiential learning tasks and activities of the practical field training that would also lend themselves both to objective assessment and experimental control. Due to limitations of the envisioned simulation environments, precision gunnery will not be part of the training, just as it is not part of the CC course itself.

While some variability in the scenarios should be present to promote learning of appropriate skills, it is unlikely that a full repertoire of scenarios suitable for a full training course will be practicable due to the level of effort of the modelling involved. Thus, several dynamic scenarios will be necessary that incorporate most of the dominant learning objectives from the CC course such as:

- Decision Making
- Situation awareness
- Crew coordination
- Own vehicle C&C
- Communications and content analysis
- Workload management
- Vigilance
- Reaction to unexpected events
- Errors (commission, omission)

2.8 Data collection

A.1.10 Recording of communications

Radio, intercom and free speech recordings will be made and analyzed to identify procedural commands and syntax errors. The audio recordings will be linked (if feasible) to video recordings to put the communications in context. Recordings made in the simulations should be relatively straight forward using DRDC equipment. Recordings made in the field may be more difficult and may not be feasible.

A.1.11 Recording of actions

Actions within the simulator will be recorded and analyzed to identify errors and correct actions based on context. Particular focus will be on inputs to the simulation in procedures and the synchrony of commands with required manual controls. Similar recordings in the field may be more difficult and may not be feasible, particularly manual control of the tank by the CC.

A.1.12 Recording of direction of gaze

The control case would be the current field training method using Leopard tanks. Some additional data collection may be attempted in the form of direction of gaze using a PC camera that may be attached to the subject's helmet provided that it does not interfere either with the safety of the subject or the training experience.

A.1.13 Recording appraisals by Directing Staff

The DSs record observations of the students using a paper form while in the field. Copies of these forms will be requested, performance details noted (anonymously) were appropriate and then the copies destroyed to protect the privacy of the subjects. Efforts will be made to ensure that there can be no attribution of data to individuals and only group statistics will be reported.

A.1.14 Measurement validity

An assessment of the experts should be done in the field prior to exposure to the simulator, however, it is recognized that this may not be practicable due to cost and availability of staff. Ideally, experts would demonstrate their proficiency in the field setting while being assessed using the same criteria as the students will be assessed during their field training session. If there has been skill fade within the expert group, the subjects should practice until they have reached the performance criteria level required of a CC. If a sufficient number of experts is available (DSs or instructors), these could form the expert group in a RTOT study rather than recruiting a separate group of qualified CCs.

The DSs and instructors who are to assess students, either in the field or in the simulator, should all assess the same sample of performances to establish a common interpretation of the assessment scheme and determine a measure of inter-rater reliability. While recording in the field may not provide adequate information to conduct a thorough assessment, recordings of performance from one of the simulators may be sufficient to allow the reviewers to have a common set of traces to evaluate as in a detailed after action review (AAR). Experts could evaluate recordings of students from a pilot study or even other experts using purpose-made simulation recordings to assess inter-evaluator repeatability as well as the state of readiness of each of the experts prior to a simulator study. Assessment of the evaluators could be done prior to a regularly scheduled course serial as the expert control group in a RTOT. This is important for the subjective assessment of subject performance to ensure uniformity of the evaluations.

2.9 Measures of performance and training effectiveness

Instructors and DS have a standard assessment sheet to record student CC performance and a standardized interpretation guide in the CC training plan (DND, 2006). Few of the assessment variables lend themselves to impartial, quantitative measures and all of the assessments are made by the instructors while observing the students conducting the traces. Some errors could be inferred from outcomes, such as intervisibility or crew safety, which could be automated to provide student feedback.

The effectiveness of the training can be judged as the number of attempts that the student makes to reach an acceptable level of performance in the various performance categories of the assessment. Communication effectiveness, as assessed by the accuracy and completeness of recording of information or providing information in the standardized format is another important metric for assessing the acquisition of skill. The computer-based training manipulations provide an opportunity to implement a capability that may represent both a pedagogical tool as well as an assessment device related to situation awareness. Endsley (1990) developed a technique called SAGAT (Situation Awareness Global Assessment Technique) that uses predetermined expert guidance to probe the state of knowledge of participants during a simulation by periodically freezing the simulation to ask questions deemed relevant by SMEs. Such an approach replicates some of the function of mentors and DSs during training.

If there is an opportunity to train to competency in all conditions and competency requires several attempts, a learning curve that describes errors committed or counts of satisfactorily completed categories is possible. If data from each trace are not available or are incomplete, the conclusions that may be drawn about the effectiveness of the training are significantly reduced because information about the rate of knowledge and skill acquisition will be missing.

2.10 Covariate measures

Covariates are variables that are often uncontrolled yet nevertheless may have a significant outcome on the results of a study and so should be measured when feasible in order to quantify their effects. For example, if CCs are renewing their qualification after a hiatus, the “Time since previously qualified” or the “Time since last conducted armoured vehicle operations” may be appropriate data to collect as they likely affect skill fade and thus influence skill re-acquisition rates.

For new CCs, variables such as “Prior exposure to armour vehicle operations”, “Prior training in other tank crew positions”, “Experience with radio communications” or “Experience with map appreciation" will be relevant. Identification of covariates is best accomplished by domain SMEs with direction from TDOs or researchers.

Observation will occur in the field when subjects are paired because 1 will be GNR for the CC, then switch, so design will have to take note of ordering effects and randomize within pairs if possible.

2.11 Time commitment

It is unlikely that a single CC course serial will provide sufficient subjects for a reliable study; however, three serials should suffice, assuming approximately 10 students per serial to be distributed randomly between the three experimental conditions (Control, Desktop Simulator, VR). More students would be advantageous provided they are approximately evenly and randomly assigned to groups; fewer students lessen the likelihood of reaching a clear inference about differences between the treatments.

If it is assumed that training in the experimental SE will proceed at approximately the same pace as training in the field, many subjects would reach proficiency after performing three simulated traces. However, students are actually exposed to three additional traces during field training when they serve as GNR while another student performs the CC role. The SEs will not be as rich a training environment, but neither will there be as many distractions, so it is difficult to predict the optimum number of traces that will be required to reach proficiency, but 10 simulated traces would be a reasonable estimate for a pilot study.

2.12 Cost effectiveness estimate

Measurement of the validity of simulation for training effectiveness may also be judged from a cost perspective, or a cost effectiveness metric. Some training approaches may not be as efficient as other approaches and thus not as effective in transferring knowledge or developing skills, but if those approaches are much cheaper to implement, then they may be desirable if cost savings are realized so that those limited training funds may be applied to additional activities. In order that the comparison is rational, costs must be made on a common basis, such as a constant level of performance achieved, or alternatively, the level of performance achieved for a fixed cost, across all training approaches. The costs for personnel and equipment are published annually in a publication called "The Cost Factor Manual" by Department of National Defence, Director Strategic Finance and Costing (DSFC 2).

The qualifications for candidate CCs are listed in QS (DND, 2006, Appendix A1). This provides a basis for estimating the opportunity cost of assigning the students to the CC training course. The QS also specifies a number of personnel and various pieces of equipment for the course. Principal among the more costly elements specified are as follows:

Personnel	**Equipment**
Course Officer (Capt/Lt)	Cartridge 105mm Blank - 320
Course Warrant Officer (WO/Adj/Sgt)	Cartridge 5.56 Blank - 1920
Administrative NCO (CPL)	7.62 Blank link - 10560
Driver (Cpl/Tpr)	Grenade Hand, Smoke: Blue, Green, Red and Yellow - 130 of each colour
Troop Leader (Capt) - 1/troop	Flare Para Comet 1260 - 350
Troop Warrant (WO/Adj) - 1/troop	Thunderflash - 200
Crew Commander Instructor (Sgt/MCpl) - 1 for every 2 students	Sim Proj Ground Burst - 420
OPFOR/Events Force Commander (Sgt/MCpl)	Grenade Hand Smoke White (TRG) - 130
OPFOR/Events Force (Cpl/Tpr) - 2/troop	5.56 C9 Service Rifle w/EIS - 1/student
	Leopard C2 Tank – 1/2 students (4 Tank min to form a troop)

3 Ethics Approval

Ethics approval is required for all human experimentation within the Department of National Defence[7] to ensure that due diligence has been performed to reduce any unusual risk imposed on the subjects, even in situations where the perceived risks are thought to be minimal. DRDC abides by the Tri-Council[8] Policy Statement on the ethical treatment of subjects, which means that a number of prescribed steps must be completed prior to conducting a study and the protocol approved by the Human Research Ethics Committee (HREC) must be substantively adhered to.

While there is no typical time required to prepare a research protocol, development and approval of a minimal risk protocol for human experimentation typically will take two months. A background literature review is typically conducted to identify risks and to ensure that the subjects are unnecessarily exposed to testing. The HREC protocol review meetings happen monthly and submissions are typically required at least two weeks in advance.

The ethics proposal details the purpose and method of the study, identifies expected risks and mitigation measures, and documents information provided to the subjects, including the content of the informed consent that is required from each subjected indicating they are aware of the risks and are willing to participate in the study.

[7] DAOD 5061-0, Research Involving Human Subjects, Issued 20 Aug 1998.
[8] http://www.pre.ethics.gc.ca/eng/policy-politique/initiatives/tcps2-eptc2/Default/

4 Experimental support

As mentioned, it is expected that the Army Learning Support Centre (ALSC) will be able to provide a significant level of programing support to develop a suitable synthetic environment for the study. This would entail further development of the Virtual Leopard 2 SE to include: the CFB Gagetown terrain; avatars that will be driven by the behavioural model, taking the place of the driver, loader and gunner; additional troop vehicles; and, OpFOR agents. Additionally, ALSC expertise may be sought to provide guidance on data collection during simulations within the SE. Technical and experimental research support will be provided by DRDC Toronto, either using in-house or contracted resources. General technical support for computer simulation is available; however, DRDC Toronto has no experience with Unity3D.

5 Funding

It is expected that CTC/ALSC has the resources required to create a SE based on the Virtual Leopard that they have previously developed. Previous meetings have explored development of the CAFB Gagetown tank training trace area as a SE in the Unity game engine.

Funding to cover development and equipment costs for the SEs that cannot be covered through GFE will be borne by the ARP 14dn Virtual Reality for Training project. This is currently envisioned to entail subject pay, a small number of VR glasses, hand controllers and computers. Large scale simulators will not be within the budget, so leveraging existing resources will be important.

Additional costs for behavioural modelling will be required from the ARP 14dn project to support the intelligent agent development and integration with the Unity SE.

6 Charter to proceed

6.1 Command support

Prior to committing to a study that may disrupt training and commit resources, approval in principle of the study should be formally established from the relevant commands, providing a charter to proceed. This would entail a letter of support and a commitment to tasks troops and equipment for the study with the expectation that reasonable efforts would be made to reduce the level of disruption. Further, to minimize costs, the study would attempt to coordinate with scheduled CC course serials.

The charter should outline expectations of the various command stake-holders for the types of deliverables resulting from the study. The charter should also outline potential problems resulting from the study, including the possibility of negative effects of training that may affect student learning, although this will be mitigated by active SME oversight from the training authority. The following **Error! Reference source not found.** lists the major suggested commitments to be agreed upon:

Commitment	OPI
Access to Crew Commander student volunteers	Armour School
Access to vehicles during field portion of CC course	Armour School
Access to shelter and power during field portion of course	Armour School
Access to Instructor volunteers	Armour School and Standards
Access to Directing Staff volunteers	Armour School and Standards
Development of the Virtual Leopard 2 SE to represent CFB Gagetown traces	ALSC
Development of Virtual Crew 3D form Avatars within the SE	ALSC
Access to Virtual Leopard 2 model	ALSC
Development of Virtual Leopard crew behavioural model	DRDC
Networking behavioural model to SE	DRDC
Staff to conduct the training effectiveness experiment	DRDC

6.2 Community support

Support for the study should be established with the training authority staff (Instructors and DSs) as well as the students to establish whether or not they are supportive and they are willing to participate. Participation must be voluntary, as outlined in the Ethics Approval section, and accommodation made for students or staffs who do not wish to be part of the study. Preliminary discussions with staff and students during a recent CC course serial suggests that the community recognizes the potential benefit of studying different training methods to determine the training effectiveness rigorously and would be open to approaches that are perceived to improve the ability of students to acquire the necessary skills to become a CC.

6.3 Client support

CDA has indicated that empirical evidence of effectiveness is an important part of the modernization process to ensure that changes to the training programme are producing the desired improvements. CDA's support is evident in their submission of this project to Director General Military Personnel Research and Analysis (DGMPRA) as part of the current CDA portfolio of research projects. Confirmation of CDA's continued support should be verified when the Charter to Proceed is ratified.

6.4 DRDC support

DRDC has included this study as a Work Breakdown Element in the Virtual Reality for Team Training project. It has been funded to 2 years; however, delays in contract procurement have consumed much of this funding window. Work has recently begun to develop and acquire some of the necessary elements for the study but more remains to be done than can be finished within the current Fiscal Year. Additional time and funding will be required to complete the study, which will require further commitment by DRDC.

6.5 Issues to be resolved

The author and the OC ALSC met to discuss a number of technical and administrative issues that need to be resolved prior to the study (see **Error! Reference source not found.**) Some of these issues will require commitment of resources and most will require commitment of time for the identified office of primary interest (OPI). Identification of the level of commitment, both in time and money, should be established prior to ratification of the Charter to Proceed.

7 Conclusion

Although there is a great deal of concern about how people learn (Lave, 1996), whether it be formal, abstract, school-house instruction or apprentice-like, on-the-job training (OJT), it seems apparent that there is a role for the spectrum of approaches to instill various aspects of the desired knowledge in the transition from a novice to an expert. Of interest to instructors and administrators alike is the most cost-effective training mix of approaches to create a cadre of experts.

Generally, the fidelity of the training environment is correlated (or proportional) to cost, but the link between fidelity and training effectiveness seems to be weak at best. As Chatham (2007), DARPA project manager responsible for adapting computer game technology for training notes, realism in the training environment should be adequate to achieve the training but not more!

A key point to remember is that VR training solutions can be useful but effective training won't happen without adequate consideration of the task, the student and the capability of the implementation. Assumptions will be made about each of these aspects during the design of the simulation and the validity of the solution needs to be measured until an adequate body of knowledge is developed that provides evidence for the adequacy of the various assumptions.

References

AGARD. (1980). Fidelity of simulation for pilot training. North Atlantic Treaty Organization (NATO), Advisory Group for Aerospace Research and Development (AGARD). AGARD-AR-159. 7 Rue Ancelle - 92200 Neuilly-Sur-Seine, France. 68.

Barab, S., Zuiker, S., Warren, S., Hickey, D., Ingram-Goble, A., Kwon, E., . . . Herring, S.C. (2007). Situationally embodied curriculum: Relating formalisms and contexts. *Science Education, 91*(5), 750-782.

Baum, D.R., Riedel, S., Hays, R.T., & Mirabella, A. (1982). Training effectiveness as a function of training device fidelity. U.S. Army Research Institute for the Behavioral and Social Sciences. AD-A140 997. Technical Report 593. Alexandria, Virginia. 65.

Bereiter, C. (1997). Situated cognition and how to overcome it. In D. Kirshner & J. A. Whitson (Eds.), *Situated cognition: Social, semiotic, and psychological perspectives* (pp. 281-300). Hillsdale, NJ: Erlbaum.

Cain, B., Magee, L., & Kersten, C. (2011). Validation of a virtual environment incorporating virtual operators for procedural learning. Defence Research and Development Canada. DRDC Toronto TR 2011-132. Toronto. 47.

Caird, J.K. (1996). *Persistent issues in the application of Virtual Environments systems to training*. Paper presented at the Third Annual Symposium on Human Interaction with Complex Systems, HICS '96. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=549502&tag=1

CDA. (2011a). Individual training and education modernization strategy: A strategic advantage for future operations. Canadian Defence Academy, Department of National Defence. 61.

CDA. (2011b). IT&E Modernization strategy. Canadian Defence Academy, Department of National Defence. 61.

Chatham, R.E. (2007). Games for training. *Communications of the ACM, 50*(7), 37-43.

Chatham, R.E. (2009). Toward a second training revolution: Promise and pitfalls of digital experiential learning. In K. A. Ericsson (Ed.), *Development of professional expertise: Toward measurement of expert performance and design of optimal learning environments* (pp. 215-246). New York, NY: Cambridge University Press.

Clark, R.E. (1994). Media will never influence learning. *Educational Technology Research and Development (ETR&D), 42*(2), 21-29.

Dahlstrom, N., Dekker, S., van Winsen, R., & Nyce, J. (2009). Fidelity and validity of simulator training. *Theoretical Issues in Ergonomic Science, 10*(4), 305-314.

Dalgarno, B., & Lee, M.J.W. (2010). What are the learning affordances of 3-D virtual environments? *British Journal of Educational Technology, 41*(1), 10-32.

de Winter, J.C.F., van Leeuwen, P.M., & Happee, R. (2012, 28-31 August). *Advantages and disadvantages of driving simulators: a discussion.* Paper presented at the Proceedings of Measuring Behavior 2012, Utrecht, The Netherlands.

de Winter, J.C.F., Wieringa, P.A., Dankelman, J., Mulder, M., van Paassen, M.M., & de Groot, S. (2012). Driving simulator fidelity and training effectiveness (pp. 9).Document.

Dede, C. (2009). Immersive interfaces for engagement and learning. *Science, 323*(January), 66-69.

DND. (2006). Tank Commander. Occupation Specialty Specification. Competency Code: AFCE. (A-P9-011-DFC/PH-B01).

DND. (2009). Qualification standard: Tank Commander. Occupational specialty specification code: AFCE. (A-P9-011-DFC/PC-B01).

Endsley, M.R. (1990). Situation awareness global assessment technique (SAGAT): Air-to-air tactical version. Northrop Corporation. Hawthorne, CA. 88.

Estock, J.L., Steizer, E.M., Alexander, A.L., & Engel, K. (2009). *Is cockpit fidelity important for effective training? Perception versus performance.* Paper presented at the Interservice/Industry Training, Simulation and Education Conference (I/ITSEC), Orlando, Florida.

Hochmitz, I., & Yuviler-Gavish, N. (2011). Physical fidelity versus cognitive fidelity training in procedural skills acquisition. *Human Factors, 53*(5), 489-501.

Ketelhut, D.J., Nelson, B.C., Clarke, J., & Dede, C. (2010). A multi-user virtual environment for building and assessing higher order inquiry skills in science. *British Journal of Educational Technology, 41*(1), 56-68.

Kirkpatrick, D.L., & Kirkpatrick, J.D. (2006). *Evaluating training programs: the four levels*.

Knerr, B.W. (2006). Current issues in the use of virtual simulations for dismounted soldier training. North Atlantic Treaty Organization (NATO), Research and Technology Organization (RTO). Meeting Proceedings RTO-MP-HFM-136. Neuilly-sur-Seine, France. 21-21 - 21-12.

Lathan, C.E., Tracey, M.R., Sebrechts, M.M., Clawson, D.M., & Higgins, G.A. (2002). Using virtual environments as training simulators: Measuring transfer. In K. M. Stanney (Ed.), *Handbook of virtual environments. Design, implementation and applications.* (pp. 403-414). Mahwah, N.J.: Lawrence Erlbaum Associates.

Lave, J. (1996). Teaching, as learning, in practice. *Mind, Culture, and Activity, 3*(3), 149-164.

Lee, E.A.L., Wong, K.W., & Fung, C.C. (2010). How does desktop virtual reality enhance learning outcomes? A structural equation modeling approach. *Computers & Education, 55*, 1424-1442.

Lee, K.M. (2004). Presence, explicated. *Communications Theory, 14*(1), 27-50.

Macedonia, M.R., & Rosenbloom, P.S. (2001). Entertainment technology and virtual environments for training and education. In R. L. Devlin & J. Meyerson (Eds.), *The Internet and the university: 2000 Forum* (pp. 79-95). Boulder, CO.: EDUCASE.

McMahan, A. (2003). Immersion, engagement and presence: A method for analyzing 3D video games. In M. J. P. Wolf & B. Perron (Eds.), *The video game theory reader* (pp. 67-86). New York, NY: Routledge.

Mestre, D.R., & Fuchs, P. (2006). Immersion et présence. Le traité de la réalité virtuelle., 10. http://www.ism.univmed.fr/mestre/pub/Pres_2005.pdf

Mikropoulos, T.A., & Natsis, A. (2011). Educational virtual environments: A ten-year review of empirical research (1999-2009). *Computers & Education, 56*, 769-780.

Muller, P., Cohn, J., Schmorrow, D., Stripling, R., Stanney, K.M., Milham, L.M., . . . Fowlkes, J.E. (2009). *The fidelity matrix: Mapping system fidelity to training outcome*. Paper presented at the Interservice/Industry Training, Simulation and Education Conference (I/ITSEC), Orlando, Florida.

Nash, E.B., Edwards, G.W., Thompson, J.A., & Barfield, W. (2000). A review of presence and performance in virtual environments. *International Journal of Human-Computer Interaction, 12*(1), 1-41.

Nichols, S., Haldane, D., & Wilson, J.R. (2000). Measurement of presence and its consequences in virtual environments. *International Journal of Human-Computer Studies, 52*, 471-491.

Noble, C. (2002). The relationship between fidelity and learning in aviation training and assessment. *Journal of Air Transportation, 7*(3), 35-54.

Patel, K., Bailenson, J.N., Hack-Jung, S., Diankov, R., & Bajcsy, R. (2006, August). *The effects of fully immersive virtual reality on the learning of physical tasks.* Paper presented at the Proceedings of the 9th Annual International Workshop on Presence, Ohio, USA.

Peli, E. (1998). The visual effects of head-mounted display (HMD) are not distinguishable from those of desk-top computer display. *Vision Research, 38*, 2053-2066.

Persky, S., Kaphingst, K.A., McCall, C., Lachance, C., Beall, A.C., & Blascovich, J. (2009). Presence relates to distinct outcomes in two virtual environments employing different learning modalities. *Cyberpsychology and Behaviour, 12*(3), 263-268.

Pollock, B., Burton, M., Kelly, J.W., Gilbert, S., & Winer, E. (2012). The right view from the wrong location: Depth perception in stereoscopic multi-user virtual environments. *IEEE Transactions on Visualization and Computer Graphics, 18*(4), 581-588.

Roscoe, S.N. (1971). Incremental Transfer Effectiveness. *Human Factors, 13*(6), 561-567.

Roscoe, S.N., & Williges, B.H. (1980). Measurement of Transfer of Training. In S. N. Roscoe (Ed.), *Aviation Psychology* (pp. 11): Ames: Iowa State University Press.

Salas, E., Bowers, C.A., & Rhodenizer, L. (2009). It is not how much you have but how you use it: Toward a rational use of simulation to support aviation training. *International Journal of Aviation Psychology, 8*(3), 197-208.

Salzman, M.C., Dede, C., & Loftin, R.B. (1999). *VR's frames of reference: A visualization techinque for mastering abstract multidimensional information*. Paper presented at the Computer-Human Interaction (CHI '99), Pittsburgh, Pennsylvania.

Salzman, M.C., Dede, C., Loftin, R.B., & Chen, J. (1999). A model for understanding how Virtual Reality aids complex conceptual learning. *Presence, 8*(3), 293-316.

Smallman, H.S., & Cook, M.B. (2010). Naive realism: Folk fallacies in the design and use of visual displays. *Topics in Cognitive Science*, 1-30.

Smallman, H.S., Cook, M.B., Manes, D.I., & Cohen, M.B. (2007, Oct 1-5). *Naive realism in terrain appreciation.* Paper presented at the Proceedings of the 51st Annual Meeting of the Human Factors and Ergonomics Society, Baltimore, MD,.

Smallman, H.S., & St. John, M. (2005). *Naïve Realism: limits of realism as a display principle.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting, Santa Monica, CA.

Smallman, H.S., St. John, M., & Cohen, H.B. (2002). Human factors of 3-D perspective displays for command and control. SPACE AND NAVAL WARFARE SYSTEMS COMMAND SAN DIEGO CA. XBSPAWARCA.

Stanton, N.A. (2006). Hierarchical Task Analysis: Developments, applications and extensions.

Stedmon, A.W., & Stone, R.J. (2001). Reviewing reality: human factors of synthetic training environments. *International Journal of Human-Computer Studies, 55*, 675-698.

Stewart, J.E., Johnson, D.M., & Howse, W.R. (2008). Fidelity requirements for Army Aviation training devices: Issues and answers. U.S. Army Research Institute for the Behavioral and Social Sciences. Research Report 1887. Arlington, Virginia. 20.

Stoffregen, T.A., Bardy, B.G., Smart, L.J., & Pagulayan, R. (2003). On the nature and evaluation of fidelity in virtual environments. In L. J. Hettinger & M. Haas (Eds.), *Virtual and adaptive environments: Applications, implications and human performance issues.* (pp. 111-128). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Stone, R.J. (2001). Virtual reality for interactive training: An industrial practitioner's viewpoint. *International Journal of Human-Computer Studies, 55*, 699-711. doi: 10.1006/ijhc.2001.0497

Watt, S.J., Akeley, K., Ernst, M.O., & Banks, M.S. (2005). Focus cues affect perceived depth. *Journal of Vision, 5*, 834–862.

Welch, R.B. (1999). How can we determine if the sense of presence affects task performance. *Presence, 8*(5), 574-577.

Willemsen, P., Gooch, A.A., Thompson, W.B., & Creem-Regehr, S.H. (2008). Effects of Stereo Viewing Conditions on Distance Perception in Virtual Environments. *Presence, 17*(1), 91-101.

Winn, W. (2003). Learning in artificial environments: Embodiment, embeddedness and dynamic adaptation. *Technology, Instruction, Cognition and Learning, 1*(1), 87-114.

Winn, W., Windschitl, M., Fruland, R., & Lee, Y. (2002). *When does immersion in a virtual environment help students construct understanding?* Paper presented at the 5th International Conference of the Learning Sciences, Seattle, Washington.

Zsambok, C.E., & Klein, G. (Eds.). (1997). *Naturalistic Decision Making*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Zyda, M. (2007). Creating a science of games. *Communications of the ACM, 50*(7), 27-29.

This page intentionally left blank.

Appendix A Outstanding issues to be resolved leading up to the Virtual Leopard 2 training effectiveness study.

Table A.1: *[Insert table caption here]* .

Topic	Comments	OPI
Authority to proceed	CA Command support for the project in the form of a charter to proceed with the work	Batty
	• This is required to prioritize CA R&D requests.	
	• Cmdt of the Armour School indicated that his School keen on participating	
	Commitment from DRDC to engage resources in the study	Cain
Select tasks and scenario	Identify a suitable experimental scenario that is manageable in the time and money that we'll have available	Batty
	• Advance to contact/attack trace	Batty
	→ having all the traces that the students have to pass in the study would make for more data and hence more sensitivity to differences	Batty
	→ assume that any terrain, weather and OpFOR 3D models would work equally well with delay, breach, etc.	Batty, Daigle
	→ need sufficient opportunity to collect data to differentiate between various conditions (for example, VR vs Workstation Simulator vs Field Training)	Batty, Cain
	Identify CC behaviours that subjects need to demonstrate and how to measure	Batty
	• Quantifiable measures of performance associated with the learning	Batty, Cain
	• focus on procedures and actions	Batty, Cain
	• CC's tasks and activities are documented in the Battle Task Standards as well as the MLPs and Turret Manual	Batty, Cain

Topic	Comments	OPI
	Contrive experimental apparatuses can accommodate the tasks and would make plausible Schoolhouse tools (i.e., Good instructional design)	Batty, Cain
Scheduling	establish some moderately hard dates/timeline	Batty, Cain
	• upcoming serials	
Conditions	Minimalist VR HMD with hand controllers as required	Batty, Cain
	A simulator built around an old LCGT or Interim Crew Gunnery Trainer	Batty, Cain
	Current field training method as control group	
	• Virtual Leopard will be employed as the immersive environment through the use of virtual reality (VR) goggles	
	• most of the SE software development will be common	
	DRDC(T) will add "avatars" to fill crew positions, so that single crew station training can be tested	Cain
	Voice Recognition capability will be added by DRDC(T) so that the avatars will be able to respond to orders	Cain
Products	Identify:	
	• What CDA wants to determine	Ball
	• What ALSC wants to determine	Batty
	• What the CA wants to know	Batty
	• What the Armour school wants to know	Batty
	Document:	Cain
	• training effectiveness	Cain
	• cost-effectiveness	Cain
	• Plausibility that VR would replace the physical simulators	Batty, Ball, Cain
	→ manoeuvring	

Topic	Comments	OPI
	→ precision gunnery	
	→ procedures	
	→ crew coordination	
	→ fireteam coordination	
	→ troop coordination	
Subjects	Availability of both Qualified and PATS/trainees	Batty
	How will we solicit troops as subjects (i.e., the paperwork required,	Batty
	• Commitment from the Armour School to supply troops	Batty
	• integrate the study with ongoing courses	Batty
Measurement	Use CC Development Rubric and Development Guide	Batty, Cain
	• Armour School SMEs to refine understanding of what gets measured and how evaluations take place.	Batty
	Identify quantitative evidence of the effectiveness	Batty, Cain
	• Time to qualify (number of traces)	
	• failure rate (reference to historical data)	
	• error rate	
	Willingness of subjects AND DSs to use head-mounted video/audio recorders during traces for subsequent data analysis and SME inter-rater reliability	Batty
	• inter-rater reliability assessment to determine differences between the DSs	Cain
	retention before retraining	n/a
	• unlikely to be able to accommodate in initial study	

Topic	Comments	OPI
	training things that cannot currently be trained	Batty, Ball, Cain
	• simple identification?	
Synthetic Environment	What are the pros and cons of VBS2 and Unity as platforms for M&S in training applications?	Daigle
	• This is important to address criticisms of study for using software that is not on the current DLSE approved list	
	• Where is VBS strong/weak in this development/application	Daigle
	• Where is Unity strong/weak in this development/application	Daigle
	DLSE commitment to VBS2 and reluctant to take on additional, competitor environments because of sunk costs	Batty, Ball, Cain
	• What are the requirements of a precision gunnery simulator that are not met in current game engines and why are these factors important for training	Batty
	• VBS2 environment is ok for what it is currently used for as a superficial training environment, but it doesn't have the horsepower to incorporate the detailed capabilities of models	Batty, Daigle
	Define simulation functionality/contents	Batty, Cain
	• What simulator development we'll need to do that training	Batty, Cain, Daigle
	→ A "virtual battlespace" (akin to VBS2 terrain) will be built by the ALSC.	Daigle
	→ ALSC will add a "rudimentary" gunnery capability and optics so that tactical training can be carried out as part of the experimentation	Daigle
	establish what capacity ALSC has to support software development in Unity needed for the scenario	Batty, Daigle

Topic	Comments	OPI
	• Assume that most of the models/behaviours could be ported over models from VBS to Unity with a little work or network both environments?	Daigle
	• Use the Unity/VBS bridge to exploit existing scenario development in VBS2?	Daigle
	• Recreate VBS2 scenario in Unity?	Daigle
	• contract out the balance of work required, possibly with oversight by ALSC	Cain, Daigle
	Establish technical POCs at ALSC for DRDC to ask integration questions	Batty, Daigle

This page intentionally left blank.

List of symbols/abbreviations/acronyms/initialisms

2D	2 dimensional
3D	3 dimensional
AAR	After Action Review
Adjt	Adjutant
ALSC	Army Learning Support Centre
C&C	Command and Control
CAF	Canadian Forces
CFB	Canadian Forces Base
Capt	Captain
CBT	Computer based training
CC	Crew Commander
CDA	Canadian Defence Academy
CER	Cost effectiveness ratio
Cpl	Corporal
CRT	Cathode Ray Tube
CTC	Combat Training Centre, CFB Gagetown, New Brunswick
DARPA	Defence Advanced Research Projects Agency
DGMPRA	Director General Military Personnel Research and Analysis
DND	Department of National Defence
DRDC	Defence Research and Development Canada
DRV	Driver
DS	Directing Staff
DSFC	Director Strategic Finance and Costing, DND
EVE	Educational Virtual Environment
FOV	Field of view
FTOT	Forward transfer of training
GNR	Gunner
h	hour
HF	Human Factors
HMD	Head mounted display
HREC	Human Research Ethics Committee
IA	Intelligent Agent
IG	Image generator
ITER	Incremental task effectiveness ratio
LCD	Liquid Crystal Display
LCGT	Land Crew Gunnery Training
LDR	Loader
LFCA TC	Land Force Command Area Training Centre
LFDTS	Land Force Doctrine and Training System
LOD	Line of departure
Lt	Lieutenant
LVCTS	Land Vehicle Crew Training System
MUVE	Multi User Virtual Environment

NCM	Non-commissioned member
NSERC	Natural Science and Engineering Research Council
OPFOR	Opposing Force
OPI	Office of primary interest
OOM	Order of March
PAT	Personnel Awaiting Training
PC	Personal Computer
PO	Performance objective
QS	Qualification Standard
RTOT	Reverse transfer of training
RW	Real world
SAGAT	Situation Awareness Global Assessment Technique
SE	Synthetic Environment
SGT	Sargent
SME	Subject matter expert
TDO	Training Development Officer
TER	Transfer effectiveness ratio
Tpr	Trooper
VCDS	Vice Chief of the Defence Staff
VE	Virtual environment
VL2	Virtual Leopard 2
VLE	Virtual Learning Environment
VR	Virtual Reality
WO	Warrant Officer

DOCUMENT CONTROL DATA

(Security markings for the title, abstract and indexing annotation must be entered when the document is Classified or Designated)

1. ORIGINATOR (The name and address of the organization preparing the document. Organizations for whom the document was prepared, e.g., Centre sponsoring a contractor's report, or tasking agency, are entered in Section 8.) Defence Research and Development Canada – Toronto 1133 Sheppard Avenue West P.O. Box 2000 Toronto, Ontario M3M 3B9	2a. SECURITY MARKING (Overall security marking of the document including special supplemental markings if applicable.) UNCLASSIFIED 2b. CONTROLLED GOODS (NON-CONTROLLED GOODS) DMC A REVIEW: GCEC DECEMBER 2012

3. TITLE (The complete document title as indicated on the title page. Its classification should be indicated by the appropriate abbreviation (S, C or U) in parentheses after the title.)

Validation of the Virtual Leopard 2 crew commander virtual reality training device : Proposed demonstration of an assessment approach

4. AUTHORS (last name, followed by initials – ranks, titles, etc., not to be used)

Cain, B.

5. DATE OF PUBLICATION (Month and year of publication of document.)	6a. NO. OF PAGES (Total containing information, including Annexes, Appendices, etc.)	6b. NO. OF REFS (Total cited in document.)
October 2014	54	61

7. DESCRIPTIVE NOTES (The category of the document, e.g., technical report, technical note or memorandum. If appropriate, enter the type of report, e.g., interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered.)

Scientific Report

8. SPONSORING ACTIVITY (The name of the department project office or laboratory sponsoring the research and development – include address.)

Defence Research and Development Canada – Toronto
1133 Sheppard Avenue West
P.O. Box 2000
Toronto, Ontario M3M 3B9

9a. PROJECT OR GRANT NO. (If appropriate, the applicable research and development project or grant number under which the document was written. Please specify whether project or grant.)	9b. CONTRACT NO. (If appropriate, the applicable number under which the document was written.)
04dn	

10a. ORIGINATOR'S DOCUMENT NUMBER (The official document number by which the document is identified by the originating activity. This number must be unique to this document.)	10b. OTHER DOCUMENT NO(s). (Any other numbers which may be assigned this document either by the originator or by the sponsor.)
DRDC Toronto TM 2013-030	

11. DOCUMENT AVAILABILITY (Any limitations on further dissemination of the document, other than those imposed by security classification.)

Unlimited

12. DOCUMENT ANNOUNCEMENT (Any limitation to the bibliographic announcement of this document. This will normally correspond to the Document Availability (11). However, where further distribution (beyond the audience specified in (11) is possible, a wider announcement audience may be selected.))

Unlimited

13. ABSTRACT (A brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), (R), or (U). It is not necessary to include here abstracts in both official languages unless the text is bilingual.)

This document outlines the proposed strategy to empirically validate the training effectiveness of a computer game-based virtual environment, Virtual Leopard 2, developed by the Canadian Armed Forces Army Learning Support Centre. The objective is to assess the training effectiveness quantitatively, comparing its use during several Crew Commander qualification courses, and make an estimate of the cost effectiveness of the use of simulation in this training application. A review of the literature indicates that empirical validation of the training effectiveness is not routinely performed despite feasible methodologies. A secondary objective of the proposed training effectiveness study is to promote the use of validation techniques for novel instructional methods to a wider audience within the Canadian Armed Forces training community. Validation exercises of this type increase confidence in the level of cost effectiveness of training provided and may be used to create a business case that has empirical support. Such studies also increase awareness of how technologies can support training as well as where they are not as effective and alternative solutions may be more appropriate.

Le présent document décrit la stratégie proposée pour valider de manière empirique l'efficacité en matière d'instruction d'un environnement virtuel basé sur un jeu informatisé (Virtual Leopard 2) créé par le Centre de soutien à l'apprentissage de l'Armée de terre (CSAAT) des Forces armées canadiennes (FAC). L'objectif est d'évaluer quantitativement l'efficacité de l'instruction offerte par cet environnement en comparant les résultats de son utilisation dans le cadre de cours de qualification pour devenir chef de char, et d'estimer la rentabilité de l'utilisation de la simulation pour cette instruction. Selon un examen de la documentation, la validation empirique de l'efficacité en matière d'instruction n'est pas effectuée régulièrement malgré des méthodes faisables. Une autre des objectifs de cette étude de l'efficacité d'instruction serait la promotion à plus grande échelle au sein de la communauté d'instruction des FAC de l'utilisation de techniques de validation pour les nouvelles méthodes d'instruction. Les exercices de validation de ce genre augmentent le niveau de confiance relativement à la rentabilité de l'instruction offerte et peuvent servir à rédiger une étude de rentabilité étayée par des données empiriques. De telles études favorisent aussi la sensibilisation sur la manière dont les technologies peuvent aider à l'instruction et démontrent les points faibles de leur utilisation en plus de présenter des solutions potentiellement plus appropriées.

14. KEYWORDS, DESCRIPTORS or IDENTIFIERS (Technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible keywords should be selected from a published thesaurus, e.g., Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus identified. If it is not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title.)

validation; virtual environment; synthetic environment; computer based training