VALIDATING USER ENGAGEMENT AND EFFECTIVENESS OF TRAINING SIMULATIONS: A mixed methods approach informed by embodied cognition and psychophysiological measures

Hiran B. Ekanayake

DSV Report Series No. 15-015
Validating User Engagement and Effectiveness of Training Simulations

A mixed methods approach informed by embodied cognition and psychophysiological measures

Hiran B. Ekanayake
Abstract

Simulation-based training has gained widespread attention recently as a response to drawbacks associated with traditional training approaches, such as high training costs (e.g. equipment), high risks (e.g. pilot training), and ethical issues (e.g. medical training), as well as a lack of availability of certain training environments (e.g. space exploration). Apart from their target training domains, many of aspects of simulations differ, such as their degree of physical realism (fidelity), scenarios (e.g. story), and pedagogical aspects (e.g. after-action reviews and collaborative learning). Among those aspects, designers have mostly focused on developing high-fidelity simulations with the expectation of increasing the effectiveness of training. However, some authors suggest that the above belief is a myth as researchers have failed to identify a linear relationship between the (physical) fidelity and training effectiveness of simulations. Most researchers have therefore evaluated the correspondence between the behaviours of trainees in both real world and simulated contexts, however, the existing methods of simulation validation using behavioural measures have a number of drawbacks, such as the fact that they do not address certain complex phenomena of skills acquisition.

Bridging the above knowledge gap, this research reports on empirical investigations using an improved methodology for validating training simulations. This research includes an investigation of the user experience of trainees, with respect to the acceptance of virtual scenarios provoking a similar psychophysiological response as in real world scenarios, and the training potential of simulations with respect to the positive transfer of training from a simulator to real world operational contexts. The most prominent features of the proposed methodology include the use of psychophysiological measures in addition to traditional behavioural measures and the use of natural (quasi-) experiments. Moreover, its conceptual framework was influenced by contemporary theories in cognitive science (e.g. constructivism and embodied cognition). The results of this research have several important theoretical and methodological implications, involving, for example, the dependency of the effectiveness of simulations on the perceived realism of trainees, which is more embodied than has been predicted by previous researchers, and the requirement of several different types/levels of adaptive training experience, depending on the type of trainee.
Keywords: training simulators, simulation validation, psychological fidelity, psychophysiological measures, embodied cognition, electroencephalography (EEG), galvanic skin response (GSR)
Sammanfattning

Träning i simulatörer har på senare år fått ökad uppmärksamhet som en respons på problem och svårigheter förknippade med traditionella träningssansatser, såsom höga kostnader (instruktörer och utrustning, etc.), hög risk (t.ex. träning av piloter), och etiska aspekter (t.ex. träning av kirurger), likaväl som avsaknaden av träningssmöjligheter och miljöer (t.ex. forskning om rymden). Bortsett från vad som specifikt tränas så skiljer sig simuleringar åt i ett flertal olika aspekter såsom fysisk realism (eng. fidelity), scenarier (handling) och pedagogiska aspekter (t.ex. genomgång efter övning och kollaborativt lärande). Bland dessa aspekter så har designerners ofta fokuserat att utveckla simuleringar med hög realism med förväntningen att detta ska göra träningen mer effektiv. Litteraturen antyder dock att denna föreställning inte stämmer och att de flesta simuleringar med hög realism inte har lyckats uppnå denna målsättning. En slutsats är därför att det finns ett behov av metoder som kan validera potentialen hos simuleringar avsedda att stödja träning – redan innan dessa används.

Enligt litteraturen så är utbildningspotentialen hos en simulering starkt kopplad till hur väl den psykologiska effekten en simulering har, stämmer överens med en verklig upplevelse. Forskning har emellertid identifierat ett flertal svagheter hos existerande ansatser för att validera simuleringar; de är ofta baserade på prestations- och/eller subjektiva mätningar; de har fokuserat en eller ett fåtal psykologiska aspekter; och de bygger på traditionella teorier. Baserat på resultat från studier av en kör-simulatör presenteras och föreslås i denna avhandling ett förbättrat ramverk för utvärdering. De mest centrala egenskaperna hos det föreslagna ramverket inbegriper användandet av psyko-fysiologiska mått tillsammans med mer traditionella mått; det konceptuella ramverket bygger på samtida teoretiska ansatser (tex konstruktivism och kroppslig kognition); samt användandet av fält (kvari-) experiment. Utöver uppnåendet av uppsatta mål för forskningen så har resultertaten ett flertal teoretiska och metodologiska implikationer. Bland dessa återfinns beroendet mellan effektiviteten hos en simulering och den upplevelse av realitet som de tränade har, vilken är mer grundläggande än vad som rapporterats i tidigare forskning, samt kravet på flera och olika typer av anpassning av träningssupplevelse för den tränade för att förhöja potentialen hos träningssimulatorer.
Nyckelord: träningssimulatorer, validering av simuleringar, psykologisk realism, psyko-fysiologiska mått, kroppsli kognition, elektroencefalografi (EEG), galvanisk hudrespon (GSR)
Acknowledgments

This research would have been another dream without the support from following people.

First of all, I wish to express my sincere thanks to my main supervisor, Professor Robert Ramberg, for believing in my ability from the beginning, guiding me in the research process, and providing feedback and advice over the years. I still remember the first advice he provided, when I met him for the first time at a SweCog conference in Vadstena some two days after I arrived in Sweden for my PhD studies. Although I wanted a lengthy discussion about my research proposal, he simply said “Hiran, you should take the responsibility for your research. We (together with Professor Tom Ziemke, who attended the same conference) will just give side support.” On the one hand this advice has given me the freedom to define my own research but on the other hand it made me realise the commitment and responsibility required for my research.

Next, I wish to thank my co-supervisor, Professor Tom Ziemke, for helping me to begin my research in Sweden, understanding my research interests, and directing me in the research process. He is the one who directed me to the InGame lab (now combined with the Interaction lab) at the University of Skövde and introduced me Professor Per Backlund. Apart from being a co-supervisor, Prof. Backlund was my closest friend in Sweden. During the hard times, especially in the winter, he arranged activities such as skiing to help me to overcome depression. I am also sincerely grateful to my co-supervisor from Sri Lanka, Professor K.P. Hewagamage, for encouraging me to start my PhD studies as soon as I completed my Master’s degree, and helping me to overcome the barriers of funding and administrative matters. I also want to thank Professor Uno Fors for being an advisor in some parts of my research.

I wish to thank my friends and the staff at the University of Skövde, especially Henrik Engström, Mikael Johanesson, Anna-Sofia Alklind Taylor, Jana Rambusch, Boris Duran, Gauss Lee, Marcus Toftedahl, Kiril Kiryazov, and Mikael Lebram. Mikael Lebram’s support was crucial when conducting experiments using the driving simulator as it was he who helped me with data collection and to overcome technical matters. The staff at the DSV, especially Dr Sirkku Männikkö Barbutiu, Fatima Santala, Brigetta Olsson, and Lars Glimbert, deserve my respect and thanks for being supportive and
friendly during my studies. My thanks also goes to the Director of UCSC, Professor G.N. Wickramanayake, former director Dr A.R. Weerasinghe, and the staff and colleagues of the University of Colombo School of Computing (UCSC) for their support during my studies.

I also want to thank the National e-Learning Centre (NeLC) project at the UCSC of the SPIDER programme of Sweden for the financial support given to my studies. Thanks also to the Swedish Cognitive Science Society (SweCog) for organising conferences, courses, and summer schools, which allowed knowledge and opinions to be shared among researchers in Sweden.

Last but not least, I thank my wife Subashinie, my mother, and my brother, for all the love and support they have given me and persevering with me during the entire period of the research.
This thesis is based on the following publications:


Contents

1 Introduction .................................................................................................................... 23
   1.1 Background to the research .............................................................................. 23
      1.1.1 Validity of driving simulators ................................................................. 24
      1.1.2 Embodied cognition ............................................................................... 25
      1.1.3 Psychophysiological measures ............................................................... 25
   1.2 Research problem and research questions ....................................................... 26
   1.3 Research approach .............................................................................................. 28
   1.4 Outline of this thesis .......................................................................................... 31
   1.5 Delimitations of scope ....................................................................................... 31
   1.6 Summary of publications .................................................................................. 32
      Paper I .................................................................................................................... 32
      Paper II .................................................................................................................. 33
      Paper III ................................................................................................................. 33
      Paper IV ............................................................................................................... 34
      Paper V ............................................................................................................... 35
      Paper VI ............................................................................................................... 35

2 Background and theory ................................................................................................. 37
   2.1 Introduction ......................................................................................................... 37
   2.2 Simulations and training ...................................................................................... 38
      2.2.1 Why use simulations for training? .......................................................... 38
      2.2.2 The requirement to validate training simulations ...................................... 39
      2.2.3 Fidelity requirement of training simulations ............................................ 40
      2.2.4 Simulations and serious games ............................................................... 41
   2.3 An educational psychological perspective of simulation based training ........... 42
      2.3.1 Different theories of learning .................................................................. 42
      2.3.2 Different assessment strategies of learning .......................................... 43
      2.3.3 Different aspects that influence the psychological processes of learning .................................................. 44
   2.4 Computational and neuroscientific methods for evaluating psychophysiological signals .................................................................................................................. 53
      2.4.1 Electroencephalography (EEG) ............................................................... 53
      2.4.2 Galvanic skin response (GSR) ................................................................. 55
   2.5 Topics in contemporary cognitive science ........................................................ 59
      2.5.1 Embodied cognition ............................................................................... 59
Appendix A: Publications ................................................................. 125
  Paper I: Determining Psychological Involvement in Multimedia Interactions ...................................................... 126
  Paper II: Game Interaction State Graphs for the Evaluation of User Engagement in Explorative and Experience-based Training Games ........ 137
  Paper III: Assessing Performance Competence in Training Games ........... 144
  Paper IV: Affective Realism of Animated Films in the Development of Simulation-Based Tutoring Systems .............................................................. 155
  Paper V: Comparing Expert Driving Behaviour in Real World and Simulator Contexts ................................................................. 170
  Paper VI: Comparing Expert and Novice Driving Behaviour in a Driving Simulator ................................................................. 185

Appendix B: Supplementary Methodology Description and Auxiliary Work ................................................................................. 203
  Appendix B.1: Supplementary methodology description for Study III .. 203
  Appendix B.2: An empirical evaluation of the Emotiv EPOC headset..... 208
  Appendix B.3: EEG and GSR correlations when watching films based on real actors based and animated characters ........................................ 224

Appendix C: Forms .............................................................................. 227
  Appendix C.1: Forms used in Study I .............................................. 228
  Appendix C.2: Consent form used in Study III .................................. 231
  Appendix C.3: Personal details form used in Study III ..................... 233
  Appendix C.4: Driving experience form 1 used in Study III ............... 235
  Appendix C.5: Driving experience form 2 used in Study III ............... 240
Figures

Figure 1. Relationship between the research problem, research questions, and related disciplines/fields/domains..................37

Figure 2. Kort’s learning spiral model..............................................................47

Figure 3. The original (left) and the current (right) models of flow state..................................................................................48

Figure 4. The 10-20 system of electrode placement.................................54

Figure 5. Three electrode placements for recording EDA.......................57

Figure 6. Graphical representation of principal EDA components (adapted from Dawson et al., 2007, p. 165)..................................................58

Figure 7. The research wheel (adapted from Rudestam & Newton, 2007, p. 5).........................................................................................70

Figure 8. The relationship between high-level objectives, publications, and the three themes of studies.....................................................79

Figure 9. Setup of experiment 1......................................................................82

Figure 10. EOG signals of each subject during (left) on-screen interaction, and, (right) off-screen interaction ...............................83

Figure 11. GSR waveforms of subjects who watched a video clip containing an exciting event.................................................................84

Figure 12. Screenshots of two types of films.................................................85

Figure 13. Setup of experiment 2....................................................................86

Figure 14. Box plots of ANOVA comparing real actor film group [1] and animated character film group [2] on the basis of (left) SCR scores, and (right) SCR amplitudes..................................................87

Figure 15. Correlation coefficients between the two groups SCR scores (solid line) and SCR amplitudes (dashed line) using a moving window of 60 seconds time frame..................................................87
Figure 16. A game interaction state graph.................................90

Figure 17. Setup of Experiment 3..................................................91

Figure 18. Relationship between the proposed quantitative method and the qualitative method of two types of driving situations over the 13 sessions..........................................................92

Figure 19. Setup of Experiment 4...................................................95

Figure 20. Screenshots of (a), the real world driving; (b), driving in the simulator; (c), OGRE based highway traffic track; (d), VDrift Monaco track; and (e), VDrift LeMans track..........................96

Figure 21. Scatterplot between means of gas (MGS) and means of speed (MSP) for different road types in the real world (i.e., Tr.11, 13, 21, and 22) and in the simulator (i.e. Tr.30, 41, and 51) for experienced/expert drivers.................................98

Figure 22. Graphical representations of the values at peaks and valleys of EEG vigilance estimators and means and standard deviations (as error bars) of the four performance measures: (a), speed; (b), steering; (c), gas; and (d), braking ............99
Tables

Table 1. Papers and research questions answered by each paper. 32

Table 2. Electrodermal measures, definitions, and typical values (adapted from Dawson et al., 2007, p. 165) .......................... 58

Table 3. General criteria for evaluation for quantitative and qualitative research ................................................................. 71

Table 4. Details of the three categories of drivers who participated in Experiment 4 ................................................................. 94
Abbreviations

ANOVA  Analysis of Variance
CAI    Computer-Aided Instruction
CNS    Central Nervous System
EBL    Experience-Based Learning
EDA    Electrodermal Activity
EEG    Electroencephalography
EOG    Electrooculography
ERP    Event-Related Potentials
fMRI   Functional Magnetic Resonance Imaging
GISG   Game Interaction State Graph
GSR    Galvanic Skin Response
HRV    Heart Rate Variability
MDO    Mentally Disordered Offenders
PNS    Peripheral Nervous System
SCL    Skin Conductance Level
SCR    Skin Conductance Response
ToT    Transfer of Training
UCSC   University of Colombo School of Computing
1 Introduction

“If you tell me, I will listen
If you show me, I will see
If you let me experience, I will learn”
(By Lao-Tse, Chinese Philosopher, 5th Century BC)

1.1 Background to the research

Over the last 30-40 years, there has been a tremendous growth in software packages for computer-aided instruction (CAI), such as e-learning, automated tutors, educational games, and simulations (Cannon-Bowers & Bowers, 2008). Apart from the many advantages of CAI, such as availability, scalability, and cost effectiveness (ibid., p. 320), it also contributes to the transformation of traditional teacher-centred instruction to student-centred learning (Moeller & Reitzes, 2011), however, early packages of CAI (e.g. e-learning) lacked certain important elements such as the ‘interactivity’ required to consider them as effective learning tools (Thomas, 2001). Today, there is a growing interest in using interactive simulations for educational and training purposes. Simulations not only enable learners to bridge experience and abstraction to deepen their understanding of learning content (see the above quote by Lao-Tse; also, Kolb (1984) on experiential learning), but also bring many other advantages for learning including unlimited repetition of educational situations and analysis of risky scenarios without endangering participants (Backlund, Engstrom, Johannesson & Lebram, 2008; SWOV, 2010). According to Drews and Bakdash (2013), simulation-based training is most effective in situations where humans are able to control or manipulate complex systems under constraints such as time pressure, safety, ethical reasons, and training costs. Such requirements are typically found in domains like aviation, health care, the military, power plants, automobiles, and refineries.

In the last few decades people have developed a number of high fidelity simulators, basically focusing on their physical realism, for various training needs, however, certain studies have shown that most high fidelity simulators have actually failed to increase the effectiveness of training, or they have had detrimental effects on training (Alessi, 2000; Bell, Kanar &
The above problem has generally been called ‘simulation validation’, which refers to the replication of simulator and real world tests to determine the extent to which measures correspond across contexts (Caird & Horrey, 2011, p. 9). Although the validation of simulations can be conducted in two forms, that is, direct (physical) and indirect (behavioural), many studies have investigated the behavioural validity of simulators considering the drawbacks of the physical validity. For instance, according to Alessi (2000), there is no linear relationship between (physical) fidelity and the training effectiveness of simulations.

Most studies that have focused on the behavioural validity of simulations have used either performance or subjective measures. For instance most existing literature on the validity of driving simulators has used performance measures such as speed, lateral position, and braking responses (Mullen, Charlton, Devlin & Bedard, 2011, p. 8) and subjective judgments and ratings given by drivers and driving instructors, however, the above measures have limitations and drawbacks (Alessi, 2000; Bell et al., 2008; Drews & Bakdash, 2013; Feinstein & Cannon, 2001; Noble, 2002). For instance, human error or inadequate performance is considered a major cause of accidents in many tasks, and is attributed to imperfect perception, insufficient attention, and inadequate information processing (Brookhuis & Waard, 2011, p. 2; Collet, Petit, Priez & Dittmar, 2005; Lal & Craig, 2001). An operator’s mental condition plays the role of a confounder (see McGwin, 2011, p. 2 for a discussion about confounding variables) if those factors have not been properly evaluated in a study which has been focused on the behavioural validity of simulations.

In an attempt to bridge the above knowledge gap, this thesis reports on research aiming at the behavioural validation of simulations informed by psychophysiological measures. This research attempts to incorporate insights about embodied cognition in order to understand the interrelationships between mind, body, and environment. For instance, the behaviour of a skilled driver during a crash avoidance situation cannot be understood based on traditional theories of cognition alone.

1.1.1 Validity of driving simulators

According to Fisher, Caird, Rizzo, and Lee (2011), driving simulators are being used by researchers in different disciplines, including psychology, engineering, and medicine, to probe several key research problems. Those problems include the efficacy of novice, commercial, and older driver training programmes; fitness to drive in patients with performance deterioration due to visual, cognitive, and motor impairments; acute and chronic effects of certain medications; the impact of alternative traffic control devices (e.g.
signs and signals) on drivers’ behaviour; and the risks and benefits of in-vehicle technologies and devices such as cell phones. There are certain dis-advantages of driving simulators, however, such as that the consequences of simulated crashes are not comparable to real crashes (Caird & Horrey, 2011, p. 4). These disadvantages ultimately raise concerns about the transferability (generalisability) of the findings of simulation studies to the open road, and therefore, simulation validation is a critical issue in simulation research.

The research reported in this thesis is basically focused on the behavioural validity of driving simulators considering the availability of facilities for conducting experiments. According to Mullen et al. (2011), behavioural validity refers to “the level of correspondence between the driving behaviours elicited in the simulator and on the roads” (p. 2). Although numerous studies have been published focusing on different aspects of driving simulation research (see Caird & Horrey, 2011, p. 12 for an analysis of publications related to driving simulations) there are only a few studies that have focused on the validation of simulations. Most of those studies were limited in scope by the fact that they were conducted under controlled laboratory conditions (see Caird & Horrey, 2011, p. 10; Engen, Lervåg & Moen, 2009 for advantages and disadvantages of different research contexts such as real world, test track, and laboratory).

1.1.2 Embodied cognition

Embodied cognition is a new perspective in cognitive science which rejects the traditional skull-bound and abstract symbolic cognition and instead puts forward the idea that cognition is typically grounded in multiple ways, including mental simulations, situated action, and bodily states (Barsalou, 2008; Wilson & Foglia, 2011).

Theories and explanations in embodied cognition are very useful for understanding several critical aspects of simulation-based training research. For instance, Noë (2004), in his proposal of enactive cognition, explained the tight coupling between perception and action by identifying perception as a kind of skilful (and intrinsically thoughtful) bodily activity on the part of the animal as a whole. These explanations help to understand the skills of an expert performer who, as Aristotle says, does “the appropriate thing, at the appropriate time, in the appropriate way” without experiencing a “representational bottleneck” (see Wilson & Foglia, 2011). Embodied cognition also justifies the use of psychophysiological measures, as discussed in the following section.

1.1.3 Psychophysiological measures

Psychophysiology is based on the assumption that “human perception, thought, emotion, and action are embodied and embedded phenomena and
measures of the processes of the brain and body contain information to understand the processes of human mind” (Cacioppo, Tassinary & Berntson, 2007, p. 4). There are other advantages of using psychophysiological measures, such as that those measures are continuous and that they appear as the only data source when the user interacts with a computer system without any explicit communication (Fairclough, 2007).

According to Kramer (2006), psychophysiological measures can be classified into those which are associated with the central nervous system (CNS) and those which are associated with the peripheral nervous system (PNS). CNS measures include electroencephalographic (EEG), event-related brain potentials (ERP), functional magnetic resonance imaging (fMRI), and electrooculography (EOG). Measures of PNS include galvanic skin response (GSR) and heart rate variability (HRV).

In simulation research, psychophysiological measures have frequently been used especially for determining the awareness level or mental workload of participants (see Nählinder, 2009), however, there is concern about whether such aspects are able to reflect true skills (i.e. the combination of cognitive, affective, and psychomotor skills) of participants. A recent successful attempt in this regard is the study of Bazanova, Mernaya & Shtark (2009) who presented a framework for evaluating certain aspects of the cognitive and psychomotor skills of musicians, using psychophysiological measures.

The main disadvantage of psychophysiological measures is the absence of a simple or a linear relationship between psychological and physiological events, which makes it difficult to index a psychological process, state, or stage with a greater accuracy (Cacioppo et al., 2007, p. 7).

1.2 Research problem and research questions

The previous section presented an important knowledge gap in the simulation-based training domain. This thesis aims to bridge this knowledge gap using the following research problem.

**Evaluate the extent to which training simulators induce behaviours similar to those in the real world from an embodied cognition perspective by using performance, psychophysiological, and subjective measures**

The relevance of behaviours for learning has been explained differently by different learning theories. For instance, behaviourism concerns only observable behaviours (e.g. physical actions), and constructivism emphasised the important role played by the human mind in knowledge construction. Theories in cognitive science, especially embodied cognition, emphasise the
multiple ways in which cognition is grounded (Chapter 2 elaborates this literature). A method which focuses on evaluating behaviours should thus incorporate both internal (e.g. cognitive, motivational, and emotional) and external (e.g. physical actions and facial expressions) aspects of behaviour. This requires different measures, that is, performance, psychophysiological, and subjective measures, for assessing those aspects, as well as a proper interconnection between the measures/aspects for describing behaviour. Considering the complexity of the above research problem, four research questions were proposed, as discussed below.

**RQ1: How can psychophysiological measures be used to determine the mental activation related to a participant’s awareness/engagement level in the simulator?**

The relevance of the above research question for the fulfillment of the research problem is argued in the following way. Many theories in education (e.g. “attention curve”) and psychology (e.g. Csikszentmihályi’s flow theory) have emphasised the importance of appropriate student engagement in a learning/training task for achieving a more productive outcome (Biggs & Tang, 2007, p. 31; Carini, Kuh & Klein, 2006; Gibbs & Habeshaw, 1992; Picard et al., 2004). Several authors in simulation research have highlighted the importance of evaluating participant perceptions of fidelity for validating simulations (Alessi, 2000; Feinstein & Cannon, 2001). According to embodied cognition, perception plays a major role in cognition and behaviour (Chapter 2). This research attempts to take various measures of engagement as a diagnostic form of perception using psychophysiological measures. This process is difficult as the researcher has to address certain challenges such as identifying appropriate sensing equipment and software tools for capturing and analysing biophysical signals under certain practical constraints (e.g. cost of equipment and signal-to-noise ratio).

**RQ2: What features of psychophysiological measures can be utilised to generalise about individuals and experimental conditions in simulator validation?**

A number of authors have reported the difficulty of generalising the results of studies based on psychophysiological measures beyond the conditions of a specific study to other settings, individuals, and outcomes (Fahrenberg, 2013, p. 323; Ravaja & Kivikangas, 2009, p. 407). This problem has been explained by Cacioppo et al. (2007, p. 7) who suggest that there is often a many-to-one relationship between psychological processes and physiological measures. As highlighted by Caird & Horrey (2011), this problem has more consequences in simulation research, as each exposure of a scene/scenario has an effect to hasten responses in subsequent exposures. This research
tackles the above problem at a fundamental level through the above research question to enable behaviours to be compared across different contexts and groups of participants.

RQ3: How do different measures of behaviour reflect the body’s role in cognition/behaviour?

Section 1.1.2 (also elaborated in Chapter 2) presented a new perspective in cognitive science called embodied cognition which is very useful for understanding human behaviour in the real world and in training simulations, however, there are only a few studies that have incorporated insights of embodied cognition into research design (e.g. Lindblom, 2007) and even fewer in simulation research. Since embodied cognition is a new perspective, most theories and methods that have been developed in traditional disembodied (mind-centred) cognitive science have become obsolete (Wilson & Golonka, 2013). This research identifies the importance of including insights into embodied cognition as well as the need to develop new methods for interpreting different measures, which is addressed through the above research question.

RQ4: How can an indirect approach, using performance, psychophysiological, and subjective measures, be utilised to validate the equivalence between simulator and real world training?

Section 1.1 discussed the drawbacks of the studies based on direct measures and the advantages of using indirect (behavioural) measures for validating training simulations (also explained in Chapter 2). Although there are at least three types of studies which are based on indirect measures (see Liu, Blickensderfer, Macchiarella & Vincenzi, 2009, p. 55), there are advantages and disadvantages to each study, such as the practical difficulties of setting up an experiment in a specific way, and ethical issues depending on the training domain. On the other hand, these studies do not suggest any specific method by which the behaviours should be evaluated. Considering these challenges, as well as the other requirements of this research (i.e. to infuse insights of embodied cognition and to use psychophysiological measures), the above research question is proposed.

1.3 Research approach

In simulation research, there are two commonly used approaches for validating training simulators: methods that involve measures of different fidelity dimensions and methods that incorporate indirect measurements (Liu, Macchiarella & Vincenzi, 2008). The former approach is based on Thorndike’s (1903) Identical Elements Theory of the Transfer of Training which
states that there would be a transfer between the first task (simulation) and the second task (real world) if the first task contained specific component activities that were held by the second task, however, this notion of validation has been challenged by several authors including Alessi (2000), indicating that there is no linear relationship between fidelity and training effectiveness of simulations. Apart from that, the identical elements approach has many disadvantages such as it is almost impossible to count the number of similar elements within the two contexts.

Conversely, the indirect measurements based approach uses behavioural measures which include performance measures, subjective judgments, and psychophysiological measures (Liu et al., 2008; Mullen et al., 2011). The performance measures help to evaluate various task specific behaviours such as vehicle speed, lane changing behaviour, and head movements in the case of a driving task, however, performance measures have limitations, such as that they are unable to capture cognition related aspects such as situation awareness (Leuchter & Urbas, 2002; Nählinder, 2009 discuss the relationship between situation awareness and task performance). Secondly, subjective judgments are the ratings given by experts through methods including questionnaires and surveys. These measures are considered much better than performance measures for some purposes, such as their usability for debriefing (Wiese, Freeman, Salter, Stelzer & Jackson, 2008, p. 288), however, subjective judgments also have limitations such as being labour and time intensive, having privacy issues, and causing interruptions to the learning experience (Drews & Bakdash, 2013; Liu et al., 2008; Picard et al., 2004). Finally, although psychophysiological measures (e.g. GSR, HRV, and EEG) offer many advantages for measuring internal aspects (e.g. cognitive, motivational, and emotional aspects) of behaviour they too have several disadvantages (discussed previously).

Although a number of studies have been conducted investigating the behavioural validity of simulations, most have focused on performance measures, with a few studies examining more complex behaviours (see Mullen et al., 2011, p. 8 for a list of performance measures used frequently in the behavioural validity of driving simulations). Most studies which have used more than one type of measure have analysed each data type in isolation from the others. For instance, Nählinder (2009) validated a flight simulator using self-ratings and psychophysiological measures; however, the results were obtained by analysing the two types of measures in isolation from each other. This research adopts a mixed methods approach (Creswell & Clark, 2006; Creswell, 2013, p. 4; Myers, 2009, p. 8) to overcome the above limitation. Creswell & Clark (2006) defined mixed methods research as follows:

“Mixed methods research is a research design with philosophical assumptions as well as methods of inquiry. As a methodology, it involves philosophical as-
sumptions that guide the direction of the collection and analysis of data and the mixture of qualitative and quantitative approaches in many phases in the research process. As a method, it focuses on collecting, analyzing, and mixing both quantitative and qualitative data in a single study or series of studies. Its central premise is that the use of quantitative and qualitative approaches in combination provides a better understanding of research problems than either approach alone.” (p. 7).

The above definition highlights the importance of philosophical assumptions in the methods selection. The philosophical assumptions of this research are influenced by several concepts, including interacting disciplines/fields/domains, research paradigms, and the researcher’s stance on research. The interacting disciplines/fields/domains include education psychology, cognitive neuroscience, computer science, and embodied cognition (see Chapter 2). For instance, educators have proposed various classification systems called ‘learning taxonomies’, such as Bloom’s taxonomy, Krathwohl’s taxonomy, and Simpson’s taxonomy (Bloom, Engelhart, Furst, Hill & Krathwohl, 1956; Krathwohl, 2002; Simpson, 1972), to use in the design and assessment of learning. Although these taxonomies are often used in academic education, they are not very effective in experience-based learning, especially when the learning constitutes different elements of cognitive, affective, and psychomotor skills, such as in driving. The view that learning can be described or assessed by observing the overt behaviours of a learner or by giving cognitive tests does not conform to the view held about embodied cognition (see Section 1.1). This research abstains from such methods in its approach, and, instead, seeks explanations that connect the mind, body, and environment. Chapter 3 presents several other philosophical assumptions that guided the research process.

This research relies on experiments, as it is interested in cause-effect relationships between independent (e.g. experimental groups and driving tasks) and dependent (e.g. vehicle speed and driver vigilance) variables, however, it preferred naturalistic (quasi-) experiments rather than controlled laboratory experiments, while minimising the effects of confounding variables (see Caird & Horrey, 2011, p. 10 for disadvantages of controlled laboratory experiments). This decision is also influenced by embodied cognition and constructivism (see Chapter 2), as the perception of a participant cannot be explained without referring to their mind, body, and situated environment. It is thus important to capture perception, as it naturally operates in a given context. In the meantime, the researcher’s deliberate involvement in the research process is also prominent in this research as an active observer when linking the high level conceptual understandings of the research problem and different results, and the behavioural patterns inferred from data and by participating in and observing the experiments (see Creswell, 2013, for explanatory and exploratory sequential mixed methods; Myers, 2009, for participant observation).
While the above description is meant to provide an overview of the research approach adopted, more specific details are presented under Chapters 2, 3, and 4, as well as in each published paper.

1.4 Outline of this thesis

This thesis is organised as follows. Chapter 1 presents an introduction to the thesis including its research problem, research questions, research approach, scope, and a summary of published work. Chapter 2 gives an overview of the theories and techniques upon which this research is based, such as educational psychology, embodied cognition, and psychophysiological techniques. Chapter 3 focuses on the methodological aspects of this thesis which includes the philosophical perspective, research process, data collection and analysis methods, and ethical considerations. Chapter 4 summarises the empirical work under three thematically organised studies. Finally, Chapter 5 presents conclusions, such as the fulfillment of research questions and problem, contributions, limitations, and future work.

1.5 Delimitations of scope

There have been many constraints and delimitations to the research conducted. While some of those delimitations have been discussed in the respective publications, this section aims to present those delimitations that are of general importance to the research as a whole.

Firstly, this research has established a boundary around its research problem by delimiting the scope of investigation on simulation-based training and focusing specifically on two aspects: trainee user experiences in simulators, and the transfer of training from simulators to real world operational contexts. Although there are many training domains, this research has another delimitation, focusing on certain advanced skills required in regular driving, such as learning to direct attention during a crash avoidance situation rather than basic learning outcomes such as manoeuvring a vehicle or special types of skills such as defensive driving. Apart from the above, the research has delimitations on its research approach by using an indirect approach and on its measures by considering only a subset of measures deemed to be reasonable (see Section 1.3 for justification).

Secondly, this research has used existing infrastructure and facilities at the university (e.g. the driving simulator, which is a mid-range driving simulator on a fixed based platform without G-forces). The research has evaluated existing simulations (modelled scenarios) rather than attempting to build new simulations.
Thirdly, although the research domain of this investigation intersects with many others, and has implications for other related domains, such as serious games, the researcher has explicitly delimited the scope of this research to avoid it deviating from its primary research domain.

1.6 Summary of publications

This section gives brief summaries of related publications, highlighting their focus, findings, and fulfillment of research questions of this thesis. Table 1 highlights the research questions answered by each paper. In-depth discussions of these publications are included in Chapter 4.

Table 1. Papers and research questions answered by each paper

<table>
<thead>
<tr>
<th>Paper</th>
<th>Research questions answered</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>RQ1, RQ2 &amp; RQ3</td>
</tr>
<tr>
<td>II</td>
<td>RQ4</td>
</tr>
<tr>
<td>III</td>
<td>RQ4</td>
</tr>
<tr>
<td>IV</td>
<td>RQ1, RQ2, RQ3 &amp; RQ4</td>
</tr>
<tr>
<td>V</td>
<td>RQ1, RQ2, RQ3 &amp; RQ4</td>
</tr>
<tr>
<td>VI</td>
<td>RQ1, RQ2, RQ3 &amp; RQ4</td>
</tr>
</tbody>
</table>

Paper I

This paper reported on an experimental study which was conducted to identify the type/level of engagement in a set of multimedia types based on psychophysiological measures.

My contribution: Literature review – 100%; Experiment design – 95%; Data collection – 100%; Data analysis – 95%; Writing and publishing – 95%

High-level knowledge contribution with respect to research questions:
*RQ1*: Psychophysiological measures were used to determine the visual attention and the depth of engagement of participants to different types of multimedia content.
*RQ2*: Phasic level changes of EOG signals and tonic level changes of GSR signals interlaced with certain scenes/events in media enabled generalised results.
*RQ3*: The successful use of psychophysiological measures demonstrated the existence of a mind-body relationship in cognition/behaviour.

**Paper II**

This paper proposed a conceptual framework for quantifying and graphically visualising the behaviours of trainees, identifying the relative significance of those behaviours for the training task.

My contribution: Literature review – 100%; Proposing the conceptual framework – 95%; Writing and publishing – 90%

High-level knowledge contribution with respect to research questions:
*RQ4*: The established relationship between performance and psychological variables through the proposed conceptual framework provides a basis for comparing simulator and real world training.

**Paper III**
This paper proposed a conceptual framework for making a numerical estimation of the performance competence of a trainee using a formula. The formula was derived based on certain relationships between performance variables described in the theories of performance motivation. The proposed conceptual framework has been validated using an experimental study.

My contribution: Literature review – 100%; Proposing the conceptual framework – 95%; Experiment design – 100%; Data collection – 80%; Data analysis – 95%; Writing and publishing – 90%

High-level knowledge contribution with respect to research questions:

**RQ4**: The established relationship between performance and psychological variables through the proposed conceptual framework provides a basis for comparing simulator and real world training.

**Paper IV**


This paper reported on an experimental study in which the aim was to investigate whether scenarios based on animated character are equally capable of triggering psychophysiological activity similar to scenarios based on real actors.

My contribution: Literature review – 90%; Experiment design – 85%; Data collection – 95%; Data analysis – 95%; Writing and publishing – 75%

High-level knowledge contribution with respect to research questions:

**RQ1**: Psychophysiological measures were used to determine the depth and nature of participant engagement in the two types of scenarios.

**RQ2**: Generalised results were obtained by averaging scores and amplitudes of phasic level GSR features.

**RQ3**: The successful use of psychophysiological measures demonstrated the existence of a mind-body relationship in cognition/behaviour.

**RQ4**: The comparison between the two types of scenarios using psychophysiological measures helped to identify a potential difference in participants’ engagement in a simulator.
Paper V

This paper reported on an experimental study comparing the real world driving behaviour of expert (experienced) drivers with a mid-range driving simulator using both performance and psychophysiological measures.

My contribution: Literature review – 100%; Experiment design – 95%; Data collection – 75%; Data analysis – 90%; Writing and publishing – 90%

High-level knowledge contribution with respect to research questions:  
*RQ1*: Psychophysiological measures were used to estimate the vigilance level of participants when driving in the two contexts.  
*RQ2*: Time-frequency based features of EEG signals interlaced with certain features of performance and subjective measures enabled generalised results.  
*RQ3*: The successful use of psychophysiological measures as well as several results that were obtained by mixing different types of behavioural measures demonstrated the existence of a mind-body relationship in cognition/behaviour.  
*RQ4*: The experiment design was based on both backward transfer and quasi-experimental studies, and it used the three types of behavioural measures.

Paper VI

This paper reported on an experimental study comparing driving behaviours of expert (experienced) and novice drivers in a driving simulator.

My contribution: Literature review – 100%; Experiment design – 95%; Data collection – 75%; Data analysis – 90%; Writing and publishing – 90%

High-level knowledge contribution with respect to research questions:  
*RQ1*: Psychophysiological measures were used to estimate the vigilance level of participants when driving in the two contexts.  
*RQ2*: Time-frequency based features of EEG signals interlaced with certain features of performance and subjective measures enabled generalised results.
RQ3: The successful use of psychophysiological measures as well as several results that were obtained by mixing different types of behavioural measures demonstrated the existence of a mind-body relationship in cognition/behaviour.

RQ4: The experiment design was based on both backward transfer and quasi-experimental studies and it used the three types of behavioural measures.
2 Background and theory

2.1 Introduction

This chapter builds the theoretical foundation upon which the research is based while extending the areas of the research problem described in Section 1.2. Figure 1 depicts the relationship between the research problem, research questions, and related disciplines/fields/domains which will be elaborated in this chapter.

![Figure 1. Relationship between the research problem, research questions, and related disciplines/fields/domains](image)
As shown in Figure 1, the research questions are bounded by the delimitations (see Section 1.5) and also by the research domain, simulation-based training. The philosophical, theoretical, and methodological foundation of this research is determined by the three interacting disciplines or fields, education psychology, cognitive neuroscience, and computer science, which are the most influential disciplines of cognitive science (see Nadel & Piattelli-Palmarini, 2003; Thagard, 2005).

2.2 Simulations and training

Section 1.1 presented a brief summary of simulation-based training while indicating certain knowledge gaps in that domain. While extending that discussion, this section focuses on developing a theoretical foundation surrounding those research issues in simulation-based training.

Simulations are intertwined with our real lives in many different ways. For instance, mental simulations help us to understand the meaning of a sentence, such as where seeing the word ‘rose’ will eventually trigger us to ‘see’ its colours, ‘smell’ its fragrance, and ‘feel’ its thorns (Barsalou, 2008; Wilson & Foglia, 2011). Mental simulations influence our decision making processes in a number of ways, including the generation of predictions, conditional probabilities, and counterfactual assessments (Moroney & Lilienthal, 2008). Apart from mental simulations, there are other external or artificial simulations which also play an important role in our real life (Gee, 2008). For instance, playing with LEGO blocks and model toys helps children to increase their experiential fidelity, and simulated sunlight is used to cure seasonal-adaptive disorder (Moroney & Lilienthal, 2008).

Today, simulations are used for a wide range of purposes, including entertainment, education, training, research, and system evaluation (Drews & Bakdash, 2013; Moroney & Lilienthal, 2008), however, according to the Modeling and Simulation Information Analysis Centre (MSIAC, 2006), the three primary domains of simulations are training, analysis, and acquisition (e.g. research and development, testing, and production). Since the scope of this thesis is limited to simulation-based training, the discussion of other uses of simulations has been omitted.

2.2.1 Why use simulations for training?

Typically, the need to use simulation for a training task emerges when the task requires humans to control or manipulate a complex system under constraints such as time pressure, safety, ethical reasons, and training costs (Drews & Bakdash, 2013). Such requirements are typically found in domains such as aviation, health care, the military, power plants, automobiles,
and refineries. For example, Boeing has deployed over 80 flight simulators around the world to meet the growing demand for over four million new commercial airline pilots\(^1\) which would otherwise be a difficult target to fulfill if traditional teaching/learning practices were involved. Other examples include the CAE 7000XR\(^2\) full-flight simulator which is being used to train pilots, technicians, and cabin crew, and the RoSS™ Robotic Surgical Simulator\(^3\) which is being used to train surgeons in minimally invasive techniques for surgery. Other advantages of using simulations include faster exposition to a wide variety of training situations, feedback from different perspectives, unlimited repetition of educational situations, automated and objectivistic assessment, demonstration of manoeuvres, analysis of risky scenarios without endangering participants, and controllable and adaptive learning tasks (Backlund, Engstrom, Johannesson, Lebram & Sjoden, 2008; ITS, 2004; SWOV, 2010).

2.2.2 The requirement to validate training simulations

Although simulation-based training has gained widespread attention in recent history, it is not completely free of disadvantages. According to Moroney & Lilienthal (2008), the disadvantages of training simulations include their failure to reflect real world performance, equipment and facility costs, attitudes of stakeholders (operators, instructors, trainees, etc.), and side effects such as cyber sickness (see Caird & Horrey, 2011 for disadvantages of driving simulators). Among those disadvantages, the most critical challenge is to ‘validate’ the positive transfer of training (ToT) of simulations. A positive transfer results when a training programme leads to increased real world performance. Other possibilities are zero transfer and negative transfer, which are not among the expectations of simulation-based training (Drews & Bakdash, 2013).

The issue with the validity of simulations emerges as a result of the limitations of time, money, and technology, as well as a lack of knowledge about which aspects of the real world need to be simulated. Designers are inclined to make assumptions about the relevance of specific factors related to model performance, which is a process that often leads to the neglect or even mis-representation of real world aspects (Moroney & Lilienthal, 2008, p. 26), however, as the ultimate goal of developing a simulation is to make it a credible model, a simulation needs to satisfy a process known as ‘verification, validation, and accreditation’ (VV&A). While accreditation is an official certification to indicate the suitability of a simulation for a particular pur-

---

1 http://www.foxbusiness.com/industries/2013/04/16/boeing-to-ramp-up-pilot-training-install-more-flight-simulators/
2 http://www.cae.com/
pose, especially supported by evidence, verification and validation try to answer the questions “Does the model accurately represent the design intent and accurately reproduce the model specifications?” and “Does the model accurately depict the ‘real’ world that it was designed to represent?” respectively (Moroney & Lilienthal, 2008, p. 28). In other words, “Validation is the process of determining that we have built the right model, whereas verification is designed to see if we have built the model right” (Pegden, Shannon & Sadowski, 1995, p. 129). Of the three processes, validation is the most important process, and can demarcate a non-effective simulation from an effective one. According to Feinstein & Cannon (2001), a process of validation should evaluate the accurate (algorithmic) representation of a desired phenomenon, participants’ perception about the phenomenon being modelled, and the phenomenon’s relationship to the real world situation being modelled.

2.2.3 Fidelity requirement of training simulations

A concept that is closely linked with validation is the ‘fidelity’ of simulations. Although there is no universally agreed definition for fidelity, according to Alessi, fidelity is “how closely a simulation imitates reality” (Liu et al., 2008, p. 62). Considering the inability to accurately conceptualise a definition for fidelity, however, researchers have been inclined to capture different fidelity dimensions. For instance, Alessi (in Liu et al., 2008) has identified six fidelity dimensions, which include equipment fidelity, environmental fidelity, psychological and cognitive fidelity, task fidelity, physical fidelity, and functional fidelity. Wyatt, Archer, and Fallows (2007) have considered only three fidelity dimensions, environmental fidelity, equipment fidelity, and psychological fidelity, however, the definitions of fidelity mainly converge into two categories, physical fidelity and psychological (or cognitive) fidelity (Feinstein & Cannon, 2001; Liu et al., 2008). Physical fidelity (also called engineering fidelity or experiential fidelity) is the degree to which the physical (sensory) characteristics of the system are replicated within the simulation. Psychological fidelity (also called action fidelity, functional fidelity, or task fidelity) is the degree to which the simulation captures the psychological aspects of the real world task or activity.

In simulation-based training literature, there are certain myths and assumptions associated with fidelity. One such myth is that high (physical) fidelity simulations always enhance training effectiveness, however, certain studies have shown that most high fidelity simulators have actually failed to increase the effectiveness of training, or else they have had detrimental effects on training (Alessi, 2000; Bell et al., 2008; Drews & Bakdash, 2013; Feinstein & Cannon, 2001; Noble, 2002). High fidelity has been shown to increase acceptance among those who are familiar with the domain (e.g. experts) as well as to increase the transfer of training. A low level of fidelity
has been recommended for novice trainees, especially during their initial stages, as high fidelity can overstimulate novice trainees while affecting them negatively. Researchers recommend an increased psychological fidelity on top of the physical fidelity of simulations to increase their training effectiveness. This has also been articulated by Alessi (2000) in a different way, who stressed the importance of trainee perceptions of fidelity rather than the actual fidelity. According to the author, depending on elements such as trainee prior experience and the complexity of the phenomenon being simulated, novice trainees perceive a medium level of fidelity as more realistic than a low or high level of fidelity.

2.2.4 Simulations and serious games

Many serious games are simulations of real world events or processes, having an educational purpose rather than being mere entertainment (Daul, 2014; Michael & Chen, 2005). The term ‘education’ refers to any desired increase in skills, knowledge, competence, and mastery, and favoured changes in attitudes, values, or behaviours (Ritterfeld, Cody & Vorderer, 2009, p. 6). Serious games are characterised as having three primary attributes: offering intrinsically motivating game play, responsiveness and immediate feedback to users, and enabling ample learning opportunities (p. 5). The authors have attributed intrinsically motivating game play to the activity similar to that of using a toy, which includes the deliberate selection of the toy, deliberate persistence of playing, and high likelihood of repetitive usage. The authors have identified three factors as the desirable outcomes of playing serious games: learning (i.e. the intentional acquisition of skills or knowledge through deliberate practice and training), development (i.e. the incidental psychological impact of game play on processes of human development such as identity or attitude formation or emotional regulation), and change (i.e. social intervention, for example, political or health behaviour). Although proponents of the serious games initiative continue to insist that they can be used as a successful learning tool, there are still certain doubts over the widespread use of serious games. For instance, embedding learning goals into games and evaluating whether those goals have been fulfilled as well as avoiding any harmful psychological and social issues associated with playing games are considered the two most important elements in serious games development (Bayliss & Schwartz, 2009; Debul, 2006; Mitchell & Savill-Smith, 2004). As a response to those opponents who argue against any learning effect of playing serious games, Gee (2009) has presented six properties of serious games as listed below:

- By choosing to play games, the players are personally attached to the goals that have been set by the designers, and, thus, they learn to solve problems by fulfilling those goals.
• Since games allow micro controlling gaming elements, the players start to feel their presence in the virtual world as a result of improved self-efficacy (this is called the embodied intimacy).
• Games facilitate experiential learning.
• Games create environments in which players have to respond in new and creative ways by finding matches between their abilities (bodies and tools) and features of the virtual world (called effectivity-affordance matches).
• Although a player’s experience in the game world is more abstract (since games are models), continued reflection on those experiences can bridge the gap between concrete experiences and more abstract and systematic understandings.
• Certain games allow players to choose between different strategies, meaning each player plays a game in different ways (called player-enacted stories or trajectories).

The above elements are equally important for simulation validation as most simulations engage learners in an entertainment experience purposefully (see Ouimet, Duffy, Simons-Morton, Brown & Fisher, 2011 for young driver problems).

2.3 An educational psychological perspective of simulation based training

The previous section (Section 2.2) discussed one of the most critical research problems in simulation-based training: the lack of effective methods for validating training simulations. According to that discussion, the effectiveness of simulation-based training is basically determined by the psychological elements rather than physical features of simulations. A question thus arises as to what extent the existing theories in educational psychology are able to describe the psychological processes associated with simulation-based training, to which this section is devoted.

2.3.1 Different theories of learning

According to Backlund et al. (2008), the learning experience in a training simulator can be best explained using the theory of experience-based learning (EBL). Other theories such as behaviourism and cognitivism assume that the world is real and external to the learner and conceptualise the objective of instruction as to map structures of the world onto the learner (Ertmer & Newby, 1993; Mergel, 1998) preferring an increased physical fidelity for simulations, however, EBL (a.k.a. constructivism or experiential learning) takes the position that knowledge cannot exist outside our minds (so it can-
not be transferred from mind-to-mind), or, in other words, what learners
know about the world stems from their own interpretations (manipulations)
of their perceptions of experiences. EBL recognises learning as a holistic
process while stressing the crucial roles played by experiences both inside
and outside the classrooms as well as the continuous reflection of learners on
their experiences. A short summary of several important studies of EBL is
given below (based on Andresen, Boud & Cohen, 1995; Fenwick, 2001).
• David Kolb presented a four stage learning cycle in which he explained
how a reflective observation of a concrete experience leads to the for-

tmation of abstract concepts (i.e. generalisations and theories), and,
which in turn, motivates the learner to test new hypothesis in future
while leading to a new learning experience;
• David Boud acknowledged that differences between individuals (inclu-
ding their past histories, learning strategies, and emotions) and between
contexts, influence the learning in different ways;
• Donald Schön, while extending the work of David Kolb, introduced three
major concepts known as double-loop learning, learning society, and re-

flection-in-action. He also recognised storytelling as a mode of reflec-
tion.

Although EBL is best when describing learning in real-life and in training
simulators (preferring a psychological fidelity), it also has certain limita-
tions. First, although it recognises the crucial role played by the learning
context, it does not explain how the context matters, as in the case of a simu-
lated context which is different from an actual operational context. Secondly,
EBL does not point out any effective and objective techniques of evaluation,
as other learning theories do.

2.3.2 Different assessment strategies of learning

In education, two types of assessment practices are used: formative and
summative (Biggs & Tang, 2007, p. 163). The purpose of formative assess-
ment is to give feedback during the learning process to allow a learner to
formulate opinions, share problem areas and reflect on how to accomplish
their goals. A summative assessment is used to grade students at the end of a
course or to accredit at the end of a programme. In simulation-based train-
ing, formative feedback is used in the form of debriefing or after-action re-
views (Wiese et al., 2008, p. 288) and summative assessment in the form of
a scoring system (Bayliss & Schwartz, 2009), however, these conventional
assessment strategies have a limited scope in simulation-based training envi-
ronments as the requirements differ from those of traditional classroom
based learning. For instance, in simulation-based training, assessment and
feedback is required to increase the motivation and self-efficacy beliefs in
addition to levelling or grading the trainees (Backlund, Engstrom, Johannesson, Lebram, et al., 2008). There is a question about the validity of
these conventional assessment strategies when they are used in simulation-based training as the skills are supposed to be applied in a different context than that in which they have been initially practiced (i.e. the problem of transfer of training).

2.3.3 Different elements that influence the psychological processes of learning

The previous two sections discussed the inability to identify a theory or an approach to effectively describe or assess simulation-based training. This section is devoted to discussing the pieces of literature that intersect with both educational psychology and simulation-based training in order to understand the psychological processes associated with simulation-based training. In other words, this section explores how the mental state of a learner can be approached and to reveal the connections between mental states and external elements such as physical actions and performance. This section is organised as follows. First, it explores technological affordances for affective learning, which range from making a simple account of affective information, through transforming an entire learning environment to a virtual one, to making generational changes over the perception and conception of learning. Next, it presents several theories that explain dynamic relationships between different emotional states, mental experiences, and states of learning while a learner is engaged with a learning task, however, as the above theories do not explain individual differences in learning, the discussion later elaborates on theories which explain certain elements deemed to precede a learner’s nature of engagement during a learning task such as why certain learners avoid challenges. Finally, it addresses another dimension which determines a learner’s nature of engagement, self-efficacy beliefs, which has a major influence in simulation-based training contexts.

Technological affordances for affective learning

The importance of emotions in learning has long been recognised by educators, for instance, Bloom and his colleagues named the affective domain as one of the three domains of learning, and Krathwohl developed a taxonomy for the affective domain (Bloom et al., 1956; Krathwohl, 2002), however, emotion evaluation in the past has been more or less dependent on qualitative or subjective measures such as questionnaires and subjective ratings which are now considered not very effective. For instance, the subjective ratings given in the form of a questionnaire usually require an interruption of the learning experience, and its validity depends on the ability and willingness of someone to articulate their own affective experience (Picard et al., 2004). Alternatives such as the use of external human observers to label affects also have weaknesses, such as their subjective biases and being la-
bour- and time-intensive. The use of technology such as a machine capable of recognising affective information has shown increased progress in recent years in recognising affect using novel techniques, such as facial and vocal expressions, chair pressure, and psychophysiological arousal (ibid.). As a result of these technological advances, researchers have been able to develop more sophisticated theories to explain the relationships between emotions and learning.

As discussed by Picard et al. (2004), technology can help elicit, sense, measure, communicate, understand, reflect upon, and respond to emotions in learning situations. For example, interactive machine models or intelligent tutoring systems can be built to perform real-time interactions with human learners to motivate, engage in and assist in new ways. One such work is the Auto Tutor project (ibid.), in which an animated agent, capable of eliciting emotions as well as responding to learner emotions, tries to hold a naturalistic conversation with a learner to help the learner to learn a subject like Newtonian physics or scientific reasoning. Extending their work, the authors (ibid.) have identified several levels according to which the relationships between technology and learning can be organised. At the basic level, computer technology can facilitate a great intensity of engagement through the power of graphics and animation as well as fostering a controllable level of challenge appropriate for the Csikszentmihályi’s flow experience or Papert’s ‘hard fun’ (these elements will be elaborated later). At the next level, the quality of engagement can be improved through a physical-digital combination of knowledge constructs (e.g. ‘programmable brick’ in LEGO Mindstorms), that is, by bringing the physical body into the learning experience (‘morphing’ the body, body-in-motion or ‘body-syntonic’). The third level is about making learning really affective, not just a simple account of affect, which is more clearly expressed in the slogan “Instead of trying to make children love the math they hate, make a math they’ll love.” This involves a greater temporal resolution of affect: from transitory emotional states, such as being happy or bored, to dispositions, such as ‘being in love,’ and an expansion of the purpose of affect: guiding the learning as well as designing learning activities so as to elicit affect in ways that will facilitate learning. Finally, with regards to the social side of affect, technology can facilitate community learning while serving the roles of emotional and inspirational mentors and fostering the creative and idiosyncratic connections to learning.

Apart from viewing the interaction between technology and learning at different levels, Sticker (2009) has elaborated one of the important social implications of affective learning: there is a big generational divide in present society depending on the dissemination of technology, and perceptions of the new digital media. He identified the victims of this disputation as ‘millennial generation learners’ (i.e. those born after 1986, also called ‘digital natives’) and the teachers of those learners (also called ‘digital immigrant’ instructors). According to Sticker (ibid.), digital immigrant instructors
want to believe that learners are same as they have always been and the same teaching/learning methods that they are familiar with should work for their students, however, there have been many changes, such as new technology and the emergence of digital media, which have enabled unprecedented forms of global communication and reach, near-instantaneous communication with others, and mobility. As conceptualised under contemporary cognitive theories, such as the extended mind thesis (Clark, 2008), the mind is becoming a hybrid of biology and technology. In other words, the mind virtually employs external symbolic material, such as digital note books and mobile phones, as their true working memory. There are therefore fundamental differences between how digital natives perceive, think, and interact, and digital immigrant instructors.

While extending his argument, Sticker (2009) elaborated on the importance of a geographically situated place for affective learning. According to the author, face-to-face affective learning in a situated place (the ‘point of observation’) is important, and fundamentally determines how knowledge is constructed (see constructivism). Social interactions and role modelling are also considered to have an impact on affective learning, as in the acquisition of new behaviours, strengthening or weakening of behavioural inhibitions, and formulating outcome expectations. Sticker (ibid.) has identified immersive virtual reality learning environments (i.e. simulations), which are designed to be experiential and intuitive, as better alternatives for the proximal real world learning situations.

**Emotions experienced during learning activities**

According to the literature, two major contributions that explain the relationships between a learning task and relevant emotions during the learning process, are the Kort’s learning spiral model (Kort, Reilly & Picard, 2001) and Csikszentmihályi’s flow experience (Csikszentmihalyi, 1997; Nakamura & Csikszentmihalyi, 2002). Kort’s learning spiral model (Figure 2) is an attempt to interweave a set of emotions relevant to learning with a set of important cognitive dynamics. The horizontal (i.e. the valence) axis represents the positive valence (or pleasant) emotions and the negative valence (or unpleasant) emotions, while the vertical (i.e. the learning) axis represents the construction of knowledge (upward) and the discarding of misconceptions (downward).
The progression of learning in Kort’s spiral model is explained as follows. A study would begin in quadrant I or II and in either case the learner's focus is on constructing or testing knowledge. In quadrant I, the learner is experiencing positive affect while being curious and fascinated about the new topic of interest. Once discrepancies start to arise between the information and the learner's knowledge structure, the learner moves to quadrant II to reduce confusion, which consists of constructive learning and negative affect such as confusion. As the learner tries to sort out the puzzle but fails, they might move into quadrant III where emotions may be negative, such as frustration, and the cognitive focus changes to eliminating certain misconceptions. After the misconceptions are discarded, the learner moves into quadrant IV. Having a fresh idea propels the learner back into quadrant I or II. Shen et al. (2009) provided some supporting evidence for Kort’s model, and some further implications. For instance, they observed engagement and confusion as the most important and frequent emotions during a learning process, and frustration as the least common. Transitions were highest between the emotional states of engagement and confusion.

The second important theory, the flow experience, is a prominent concept in psychology, proposed by Mihaly Csikszentmihalyi (Nakamura & Csikszentmihalyi, 2002). Csikszentmihalyi described flow as “being completely involved in an activity for its own sake. The ego falls away. Time flies. Every action, movement, and thought follows inevitably from the previous one, like playing jazz. Your whole being is involved, and you're using your skills to the utmost” (in Geirland, 1996). The flow experience is regarded as better than happiness, which experience usually depends on task-unrelated and favourable external conditions, such as a warm sunshine. The
flow experience is related to the task, however, and it enhances a person’s cognitive and related processes, including intrinsic motivation, attention and concentration, to optimal levels. Flow is conditioned on certain factors such as clear proximal goals (rules) and immediate feedback about progress, a balance between the perceived challenge of the task and the perceived skills of the performer (i.e. an activity that is neither too easy nor too difficult), the activity should be favourable or intrinsically rewarding to the person (e.g. gardening or driving) and cause a lack of awareness of bodily needs. Figure 3 shows two graphs related to flow: Csíkszentmihályi’s original model of flow and the new flow model proposed by the Milan group.

Csíkszentmihályi’s original model of flow predicts three regions: a flow channel in which task challenge matches skills, a region of boredom in which opportunities for action relative to skills have declined, and a region of anxiety in which a task challenge has increasingly exceeded the capacities for action. According to the new model, flow can occur when both skills and task challenges exceed average levels, or in other words: “… when individuals perceive greater opportunities for action than they encounter on average in their daily lives, and have skills adequate to engage them” (Nakamura & Csikszentmihalyi, 2002, p.95). The new model has proposed a fourth state, apathy, which corresponds to low challenges and low skills. Apart from the above changes, the new model has divided the challenge/skill terrain into eight experiential channels and a series of concentric rings to represent increasing intensity of experience. The frequency and quality of flow experience varies among individuals, however, and Csikszentmihályi has recognised a certain personality that tends to enjoy life called the ‘autotelic personality’. The distinguishing meta-skills or competencies of autotelic personality are a general curiosity and interest in life, persistence and low self-centredness, which motivates a person to cultivate an openness to intrinsic rewards (Nakamura & Csikszentmihalyi, 2002).
Motivation for learning

Although there is a general opinion that student engagement is critical for a successful learning outcome, according to certain educators (in Carini et al., 2006), the relationship between student engagement and performance is complex. In other words, successful student engagement in a learning task is not something which can just be determined by observing a learner, but something which concerns mental states, including the internal motivation of learners.

As suggested by Biggs and Tang (Biggs & Tang, 2007, p. 34), there are four types of motivation applicable to students: extrinsic, social, achievement, and intrinsic. Extrinsic motivation depends on the value or importance students attach to the outcomes (i.e. consequences) of the task, such as material rewards for success or punishments for failure. The quality of learning under extrinsic conditions is believed to be low, however, as negative reinforcements such as fear of failure can easily demotivate students. Social motivation occurs when students learn in order to please people who are important to them, such as their parents. These students sometimes adopt models, such as university teachers. If a student is motivated in order to enhance their ego by competing with other students to achieve a higher graded performance, it is due to achievement motivation. Although achievement motivation can lead those motivated students to deep learning, it is not good in general, as not all students are motivated under competitive conditions and it does not facilitate collaboration. Finally, intrinsic motivation occurs when a student is interested in the task or activity itself for intellectual pleasure, independently of any rewards. Intrinsic motivation is considered ideal for learning.

Although Biggs and Tang (ibid.) have suggested a simple relationship between motivational constructs and performance, according to motivation theorists such as Hidi and Harackiewicz (2000) and Elliot and Church (1997), the relationship between the two elements is very complex, as elaborated below.

A student’s ability and effort are considered the primary determinants of performance. Whilst there is very little that educators can do about student ability, effort is believed to be dependent on several factors, such as interest and motivation, which are externally controllable to some extent. Interest is conceptualised as both a psychological state (i.e. a state characterised by focused attention, persistence effort, etc.) and the disposition (i.e. a habitual affective way) of a person, which emerges as a result of an interaction between an individual and certain elements (e.g. objects, events, ideas) of the situated environment. There are two types of interests: situational and individual. Individual interest (also called personal interest) is a relatively persistent motivational orientation or personal disposition to attend to or engage in certain tasks, and it is typically associated with positive feelings and in-
creased cognitive functions including attention, recognition, and memory. Individual interest is thus considered an important determinant of motivation and learning. Conversely, situational interest is a spontaneous interest that is triggered by certain conditions, situations or stimuli in the environment such as novelty, violence, and uncertainty. Situational interest is characterised by a focused attention and an immediate affective reaction with a broader range of emotions, which may or may not last for a prolonged period. Since situational interest offers a way to motivate otherwise academically unmotivated students by concentrating on the features of the task and the learning environment (which includes social factors such as working in the presence of others), educators consider it an important motivational construct in student learning (Hidi & Harackiewicz, 2000). For instance, a boring or an uninteresting task can be changed into a more interesting task by making games out of it and helping students to intentionally regulate their interest in the task. One challenge of situational interest is to maintain it, so that it becomes self-determined, autonomous, and enjoyable (i.e. a wide range of knowledge patterns will be activated making connections and developing new hypotheses etc.). This process, the affective-cognitive synthesis (or undivided interest), can ultimately contribute to the development of individual interest and intrinsic motivation. Individual interest can also play a role in situational interest to moderate the impact of environmental factors.

Although many researchers use the terms ‘interest’ and ‘intrinsic motivation’ interchangeably, individual interest is viewed as a precondition to intrinsic motivation. Situational interest can also correspond to intrinsic motivation, if it can be maintained. Although early studies suggested that external interventions (e.g. rewards, competitions, and deadlines) could undermine intrinsic motivation, recent studies have shown that it depends on the complexity of the activity and the length of involvement. For instance, both intrinsic rewards, which are inherent in interesting activities, and extrinsic rewards, such as performance feedback, could improve task engagement while contributing to the emergence of individual interest and intrinsic motivation. Another theory of motivation, the relative autonomy theory (Hidi & Harackiewicz, 2000), provides a more useful way to conceptualise the motivational basis of learning than viewing it on intrinsic or extrinsic grounds. According to the theory, relative autonomy is determined by the degree of self-determination of an individual and the two processes called internalisation and integration, which allow an individual to assimilate external factors themselves.

Further advances in motivation theories have confirmed that viewing motivation as an intrinsic-extrinsic continuum is not sufficient, and that such an evaluation should consider how both internal and external factors might work together to facilitate motivation and learning. This requires a clear understanding of how interests interact with personal goals. Personal goals are considered as behavioural intentions to guide the thoughts, feelings, and
behaviour of individuals in specific contexts. For instance, achievement goals represent an individual’s motivation for task engagement, especially in an academic context. There are two types of achievement goals, called ‘mastery’ (or learning) goals and ‘performance’ (or ego) goals. Mastery goals are presumed to orient people towards mastering tasks and the development of competence, as well as “… persistence in the face of difficulty or failure, the achievement of self-referenced standards, and the recognition that effort and risk-taking are elements of success” (Hidi & Harackiewicz, 2000, p.160). In contrast, performance goals lead individuals to the demonstration of competence relative to others, or to outperform them, however, they can have negative effects on learning, such as the withdrawal of effort in the face of failure, avoiding challenge, and a preference for strategies to minimise effort. Two variants of performance goals are performance-approach (or relative ability) goals and performance-avoidance (or extrinsic) goals (ibid.). Performance-approach goals are focused more on attaining competence and favourable judgments, whereas, performance-avoidance goals are focused on consequences, or avoiding unfavourable judgments of competence. Students who have adopted mastery and performance-approach goals have shown higher levels of self-regulation on potential positive outcomes (i.e. task mastery and normative competence) while promoting affective-cognitive processes which are characterised by excitement and task absorption, and producing mastery pattern of achievement outcomes. Performance-avoidance goals are characterised as self-regulation according to potential negative outcomes of affective-cognitive processes which are characterised by anxiety and task distraction, and helpless patterns of achievement outcomes.

Based on a proposed hierarchical model of approach and avoidance achievement motivation, Elliot and Church (1997), using an empirical approach, explained a set of possible relationships between motivational dispositions (i.e. achievement motivation and fear of failure), achievement goals (i.e. mastery, performance-approach, and performance-avoidance), task-specific competence expectancies, and achievement relevant outcomes/behaviour (i.e. intrinsic motivation and graded performance). According to the authors, mastery goals were grounded in achievement motivation and high competence expectancies; performance-approach goals in both motivational dispositions and high competence expectancies; and performance-avoidance goals in the fear of failure and low competence expectancies. They found that mastery goals facilitated intrinsic motivation, performance-approach goals enhanced graded performance, and performance-avoidance goals undermined both intrinsic motivation and graded performance. There are other types of relationships proposed by researchers, however, such as individual interest as a predictor of goal adoption; individual interest as an outcome of goal adoption; the relationship between mastery goals and interest as reciprocal (not unidirectional); and individual interest as a precursor of mastery goals and situational interest as a consequence of
mastery goals (Hidi & Harackiewicz, 2000). Hidi and Harackiewicz (2000) emphasised the importance of considering a breadth of influences such as motivational processes (i.e. competence valuation and task involvement), an individual’s other goals, personality, and contextual factors for predicting goal adoption and their relationships to motivational variables.

**Self-efficacy beliefs during learning**

Albert Bandura (1993) explained how perceived self-efficacy influences cognitive, motivational, affective, and select processes as follows. Self-efficacy beliefs are more central and pervasive than any other mechanism of personal agency. Among the many relationships, self-efficacy beliefs can be discussed in relation to personal goals, skill utilisation, social comparison and feedback, and views about the controllability of the environment.

A person who has a high sense of efficacy is more likely to set challenging goals, as well as to have a stronger commitment to their goals. Since such individuals situate self-satisfaction conditions in goal fulfillment and discontent with substandard performance, their effort remains until goal accomplishment. Furthermore, they attribute their failures to insufficient effort and consider difficult tasks as challenges to be mastered rather than as threats to be avoided. In contrast, a person who doubts their efficacy tries to avoid difficult tasks, seeing them as threats. They concentrate on the many things that can go wrong and give up quickly, or recover slowly in the face of difficulties or failure.

Self-efficacy beliefs also have an influence on human abilities. Human ability is conceptualised in different ways by different people, such as that ability is an acquirable skill that can be increased by gaining knowledge and competence; ability is an inherent capacity, with performance as its diagnostic form; and that ability is a biologically shrinking capacity, with increasing age and faulty performance as indications of declining capacity. Nevertheless, the above conceptions have implications on thought processes and performance attainments, or, in other words, a person may perform poorly, adequately or extraordinarily, depending on their level of self-efficacy beliefs, despite how much the person is knowledgeable or skilful.

Social comparison and feedback also have implications for self-efficacy. For example, seeing oneself surpassed by others or highlighting deficiencies in feedback can have negative effects on self-efficacy, analytical thinking, and performance. On the other hand, seeing oneself gain progressive mastery or asserting the gains during feedback have positive effects on those processes and outcomes. It is therefore suggested that learning environments that consider ability as an acquirable skill may avoid competitive social comparison, and, instead, promote self-comparison of progress and personal accomplishment. People’s efficacy beliefs about the controllability of their environment have two components: producing changes through persistent effort and the creative use of abilities and resources, and the modifiability of
the environment. In this endeavour, people with firm beliefs gain the most advantage, even in environments with limited opportunities and many constraints.

People’s self-efficacy beliefs have implications for how much stress and depression they experience in difficult or threatening situations, and their level of motivation. It is believed that both perceived coping self-efficacy and thought control helps to reduce anxiety arousal and avoidance behaviour. A decreased coping efficacy can have negative physiological effects, such as accelerated heart rate, increased blood pressure, and the activation of stress-related hormones, all of which should be progressively overcome by guided mastery experiences. Thought control is necessary to turn off disturbing thoughts that increase stress. Finally, personal efficacy beliefs can shape people’s life paths by influencing their choice of activities and environments: for instance, they can select activities and situations they believe they can cope with.

2.4 Computational and neuroscientific methods for evaluating psychophysiological signals

Section 2.2 emphasised the importance of considering the psychological elements of simulations for an increased transfer of training, however, as discussed in Section 2.3, the existing methods in educational psychology are unable to capture psychological elements of learners in reliable and effective ways. This section thus aims to explore methods and techniques in neuroscience and computation which can be used to make objective measures of psychological elements. Although there are many such methods (see Hsieh, 2003; Savoy, 2001; Thagard, 2005), this section focuses only on a few that are practically feasible in terms of cost, safety, and flexibility, but still effective in detecting the mental state of learners.

2.4.1 Electroencephalography (EEG)

Electroencephalography (EEG) can be considered the most effective brain mapping technique in practical terms, including safety, cost, and flexibility. It is used for a wide spectrum of applications in addition to clinical settings, such as biofeedback therapy and computer games.

EEG signals mainly constitute currents which flow during synaptic excitations of the dendrites of many pyramidal neurons in the cerebral cortex (Seeley, Stephens & Tate, 2008, p. 382; Tatum, Husain, Benbadis & Kaplan, 2008; Teplan, 2002). Since this thesis is focused on using EEG for detecting mental states, only the elements and requirements related to the recording of EEG potentials are elaborated here (an interested reader could refer to Seeley
et al. (2008) and Thagard (2005) to gain knowledge about the functional
correspondence between brain potential and the cognitive/emotional
activities of the brain).

Scalp potential (EEG) involves sinusoidal waveforms and reflects the
pooled synchronous activity of a large population of neurons. The amplitude
of EEG potential usually varies between 0.5 to 100 µV, and there are several
dominant waveforms in an EEG recording, called delta (0.5-4 Hz), theta (4-8
Hz), alpha (8-12 Hz), beta (13-30 Hz), and gamma (30-45 Hz) waveforms. A
typical EEG recording system contains a set of electrodes with conductive
media, an amplifier with filters, an analogue-to-digital converter, and a re-
cording device. Recording using a minimal mono-channel EEG system re-
quires one active electrode, one (or two linked) reference electrodes, and one
ground electrode (Teplan, 2002), however, for a clinical study a minimum of
21 electrodes are required with an electrode impedance of 100-5000 ohms
(Tatum et al., 2008).

There are standardised electrode placement systems, such as the interna-
tional 10-20 system or modified combinational system with 10-10 system
(Tatum et al., 2008; Teplan, 2002). Figure 4 shows the electrode placement
of the original 10-20 system in which the head is divided into proportional
distances by the intervals of 10% to 20% from the skull landmarks (i.e. nasi-
on, preauricular points, and inion).

![Figure 4. The 10-20 system of electrode placement](image)

In the above electrode placement system (Figure 4), different labels have
been used to name the different sites on the scalp. For example, FP is for
frontopolar, F for frontal, T for temporal, O for occipital, C for central, and P
for parietal. In addition, the number or letter following the label reflects the
left (odd number), midline (“z”) or right (even number) part of the brain.
Each of these positions in the system corresponds to a specific functional
localisation in the brain. For example, F7 corresponds to rational activities,
Fz to intentional and motivational centres, F8 to emotional impulses, C3, C4 and Cz to sensory and motor functions, P3, P4 and Pz to perception and differentiation, T3 and T4 to emotional processes, T5 and T6 to memory functions, and O1 and O2 to primary visual functions (Teplan, 2002).

In addition to the electrode placement, EEG configurations differ from each other depending on the way the pairs of electrodes are configured, called a montage (Tatum et al., 2008; Teplan, 2002). According to the literature, the most commonly used montages are the common reference and bipolar montages. In the common reference montage each channel uses the same reference electrode, such as midline (e.g. Cz), linked-ears, linked-mastoids, ipsilateral-ear, contralateral-ear, C7, or the tip of the nose. On the other hand, the bipolar reference is arranged either longitudinally (called “double banana”), as sequentially linked in straight lines from front to the back of the head, or in a circumferential pattern. This can signify the absolute electrographic sites of maximal negativity (or positivity) by phase reversals. Apart from different arrangements of electrode placement, EEG potentials require an amplification stage capable of selectively amplifying the signals. According to Teplan (2002), an EEG amplifier must satisfy the following requirements:

- The physiological process to be monitored should not be influenced in any way by the amplifier
- The measured signal should not be distorted
- The amplifier should provide the best possible separation of signal and interferences (such as from a 50/60 Hz power line, tissue/electrode interface and noise)
- The amplifier has to offer protection for the patient against any danger of electric shock
- The amplifier itself has to be protected against damage that might result from high input voltages.

The recommended gain requirement of the amplifier is set as 100-100,000 and a high common-mode rejection ratio (at least 100 dB) with a high input impedance (at least 100 Mega Ohms) is recommended to reduce the noise impact of the amplifier (Teplan, 2002). Additionally, at least a 12 bit analogue-to-digital converter with an accuracy higher than the overall noise (0.3-2 µV) and a sampling frequency of 128-1024 Hz is recommended for the EEG recording system. Other than that, analogue and digital filters are used for reducing undesired bio potentials. For instance, a high-pass filter is used to reduce the bioelectric flowing potentials generated from breathing.

2.4.2 Galvanic skin response (GSR)

In addition to brain mapping techniques, there are other biophysical techniques for measuring psychophysiological activity of humans such as galvanic skin response (GSR) and heart rate variability (HRV). Of these, the
GSR, formally known as electrodermal activity (EDA), is a widely used response system for assessing psychological states or processes, such as attention, information processing, and emotion, based on changes in skin conductance or resistance (Dawson, Schell & Filion, 2007, p. 159). The primary reason behind the popularity of EDA is the simple and linear relationship between EDA and the level of autonomic arousal: skin conductance increases with the increase of autonomic arousal. A typical EDA recording has two types of waveforms: skin conductance level (SCL) and skin conductance response (SCR) (ibid.). The SCL is the slowly changing level of an EDA recording and it is associated with tonic phenomena or the general changes in autonomic arousal, whereas the faster changing elements, or phasic increases in conductance on top of the tonic level, are known as SCR. The SCRs are generally associated with external stimuli or emotion evoking events.

Although it was previously believed that EDA is associated with the activation of both sympathetic and parasympathetic branches of the autonomic nervous system, according to recent findings EDA is associated only with the sympathetic activation (Dawson et al., 2007, p. 161). More specifically, EDA is a result of the activity of palmar sweat glands (as discussed in secretory theory) with the purpose of grasping rather than evaporative cooling, however, the neural mechanisms for excitatory and inhibitory influences on sympathetic activation are said to be very complex, for example, in the contralateral-cortical and basal ganglion influences from premotor cortex and frontal cortex, hypothalamus and limbic system including amygdala and hippocampus, reticular formation in the brainstem, and the activation of brain areas involved in evaluating stimulus significance, including ventromedial prefrontal cortex, right inferior parietal region, and anterior cingulate. According to Dawson et al. (2007, p. 168), EDA is influenced by the activation of the neurophysiological behavioural inhibition system (the anxiety system) involved in responding to punishment, passive avoidance, and frustrated non-reward. There is another paradoxical response associated with EDA depending on the psychological state of an individual (Shepherd, 2008), which is that, depending on the psychological states of being paratelic (active) or telic (thinking), an individual may feel high arousal as pleasant or unpleasant, respectively.

According to Dawson et al. (2007, p. 159), there are two methods for recording EDA, known as the exosomatic method and the endosomatic method. The exosomatic method, first discovered by Fere in 1888 and the most frequently used method, consists of passing an external current across the skin, and the endosomatic method (also called the Tarchanoff method) measures EDA activity without involving an external current. Apart from the EDA recording system, other considerations are the type of recording electrodes (e.g. silver), electrode paste (e.g. silver chloride), electrode placement...
(Figure 5), and general environmental considerations (e.g. temperature, humidity, and time of day).

![Diagram of electrode placements](image)

**Figure 5. Three electrode placements for recording EDA**

*Note: Placement #1 involves the volar surfaces on medial phalanges, placement #2 involves the volar surfaces of distal phalanges, and placement #3 involves the thenar and hypothenar eminences of palms (adapted from Dawson et al., 2007, p. 163).*

The two types of EDA components, tonic SCL and phasic SCR, have different characteristics depending on the recording conditions (Dawson et al., 2007, p. 164). For instance, tonic SCLs are sensitive to stimulus novelty, intensity, and significance. Tonic SCLs thus vary widely across different subjects, as well as depending on different psychological states of the same subject. A phasic SCR can occur in relation to an identifiable stimulus (called “specific” SCR) or in the absence of an identifiable stimulus (called “spontaneous” or “non-specific” SCR). Figure 6 shows the components of a typical SCR waveform and Table 2 gives a list of measures, definitions, and the typical values of EDA.
Table 2. Electrodermal measures, definitions, and typical values (adapted from Dawson et al., 2007, p. 165)

<table>
<thead>
<tr>
<th></th>
<th>Description</th>
<th>Typical Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency of NS-SCRs</td>
<td>Number of SCRs in absence of identifiable eliciting stimulus</td>
<td>1-3 per min</td>
</tr>
<tr>
<td>SCR amplitude</td>
<td>Phasic increase in conductance shortly following stimulus onset</td>
<td>0.1-1.0 uS</td>
</tr>
<tr>
<td>SCR latency</td>
<td>Temporal interval between stimulus onset and SCR initiation</td>
<td>1-3 s</td>
</tr>
<tr>
<td>SCR rise time</td>
<td>Temporal interval between SCR initiation and SCR peak</td>
<td>1-3 s</td>
</tr>
<tr>
<td>SCR half recovery</td>
<td>Temporal interval between SCR peak and point of 50% recovery of SCR amplitude</td>
<td>2-10 s</td>
</tr>
<tr>
<td>SCR habituation</td>
<td>Number of stimulus presentations before two or three trials with no response</td>
<td>2-8 stimulus presentations</td>
</tr>
</tbody>
</table>

Key: SCL, skin conductance level; SCR, skin conductance response; NS-SCR, nonspecific skin conductance response

The EDA imposes certain challenges when it is used in psychophysiological research (Dawson et al., 2007, p. 167). For instance, the occurrence of another SCR before the previous SCR has fully recovered; different ranges of SCL due to individual differences of psychological and physiological phenomena; and SCR habituation effects (i.e. decline of SCR amplitude with the repetition of eliciting stimulus); and the occurrence of non-specific SCRs.

Figure 6. Graphical representation of principal EDA components (adapted from Dawson et al., 2007, p. 165)
2.5 Topics in contemporary cognitive science

Cognitive science can be broadly defined as the interdisciplinary scientific study of minds, brains, and processes such as perception, attention, memory, language, and intelligence in humans, animals, or machines (Nadel & Piattelli-Palmarini, 2003; Thagard, 2005). It has a very long history, which can be traced back to ancient Greek philosophers in the western scientific tradition, and there are associations with eastern religious/philosophical traditions such as Hinduism and Buddhism.

According to the literature, cognitive science can be classified into two major streams: traditional and contemporary. Traditional cognitive science (a.k.a. cognitivism) explains cognition in terms of sequential series of processing stages parallel to a digital computer, including perception, attention, short-term memory, rehearsal, and long-term memory (Eysenck & Keane, 2005; Galotti, 2004; Kahneman, 1973; Miller, 1956; Moore & Egeth, 1998; Nadel & Piattelli-Palmarini, 2003; Seeley et al., 2008; Thagard, 2005), however, this stream has mistaken and neglected crucial elements of thinking, including consciousness, emotions, and the body’s involvement in cognition (Thagard, 2005, p. 140). Conversely, contemporary cognitive science (a.k.a. embodied cognitive science) explains cognition with reference to the crucial roles played by the biological body, including modal systems for perception, action, and introspection. This thesis identifies the importance of this contemporary cognitive scientific perspective from several directions. For instance, traditional cognitive science cannot explain the skilful actions of experts including the tight coupling between perception and actions using its representation oriented cognitive architecture. It is very difficult to understand/explain the behaviour of different categories of drivers in a simulator based on the theories of traditional cognitive science. Finally, only embodied cognitive science is able to make meaningful connections between an agent’s physiology, and their perception and actions in a given context, which are crucial for the scope of this research.

2.5.1 Embodied cognition

According to Nadel and Piattelli-Palmarini (2003), biology and cognitive science had separate paths of development for a considerable period after the 1950s. During that period, cognitive science primarily relied on a symbolic framework that involved representational structures and the computational procedures that operate on those structures. According to this model, knowledge resides in a semantic memory system in the form of representation called amodal symbols, which are abstract representational structures
independent of the brain’s modal systems, for perception (e.g. vision), action (e.g. movement), and introspection (e.g. mental states) (Barsalou, 2008; Wilson & Foglia, 2011), however, this view has been challenged by contemporary cognitive scientists in various ways, including that there is little empirical evidence supporting the involvement of amodal symbols in cognition, that traditional theories fail to explain how cognition interferes with perception and action, and that no neurological evidence supports the existence of amodal representations and supporting processes in the brain (Barsalou, 2008). A recent attempt to rectify the unsuccessful path taken by cognitive science is called ‘grounded cognition’, which rejects traditional amodal representations and instead assumes that cognition is typically grounded in multiple ways, including simulations, situated action, and bodily states (Barsalou, 2008). Although phrase embodied cognition is often used to refer to the same literature as that about grounded cognition, as noted by Barsalou, most theorists are mistakenly think that it stresses only the necessity of bodily states in cognition: cognition cannot proceed without depending on bodily states. As will be seen later in this section, however, the different ways that cognition is grounded all depend on the features of the physical body of an agent: cognition is indeed embodied. Considering this dependence, theorists have proposed three roles or functions that the body plays in cognition: the body as a constraint on cognition, the body as a distributor of cognitive processing, and the body as a regulator of cognitive activity (Wilson & Foglia, 2011).

Although the embodied perspective of cognition is new to the cognitive science community, its roots can be found throughout the history of cognitive science (Barsalou, 2008; Wilson, 2002; Wilson & Foglia, 2011). For example, Jean Piaget’s work on developmental psychology emphasised the role of biological processes, as well as environmental experience, in the development of cognitive abilities. Wilson and Foglia (2011) presented the early works that were influential in grounded cognition under two versions of cognitive science: the narrow sense and the broad sense. The narrow sense refers to the traditional central cognitive processing, and the broad sense to the full range of perceptual, cognitive, and motor capabilities. Those early works on the narrow sense included findings related to metaphors in languages, enactive cognition, and Rodney Brook’s embodied approach to robotics. George Lakoff and Mark Johnson’s 1980s book, “Metaphors We Live By”, presented a discussion of metaphors and their relevance for cognition: the metaphors embedded in figurative languages actively structure much of cognition. For example, many central cognitive processes, such as those concerning space (i.e. concepts such as “front”, “back”, “up”, and “down” which are articulated in terms of a body’s position in, and movement through, space) and time, were expressed and influenced by metaphors. In their book titled “The Embodied Mind”, Varela et al. (1991) attempted to bring Merleau-Ponty’s phenomenological perspective and Buddhist philoso-
phy to cognitive science to emphasise an agent’s mutual interactions between its physiology, its sensorimotor circuit, and the environment. According to their theory, only a creature with certain features (e.g. eyes, hands, legs, and skills) can possess certain kinds of cognitive capabilities (e.g. colour experience and categorisation). Finally, as exemplified in Andy Clark’s 1997 book, titled “Being There: Putting Mind, World, and Body Back Together”, Brook’s subsumption architecture was successful in implementing computational intelligence through a bottom-up approach while avoiding painful and complicated internal algorithms and representations. An important implication of his work is that minds are not for just thinking (as traditionally conceived) but also for doing or acting.

That early work in the broad sense included J.J. Gibson’s ecological approach to perception, application of dynamical systems theory to developmental psychology (as presented in the work of Esther Thelen and Linda Smith in 1994), and phenomenology. Gibson’s ecological approach to perception challenged the idea that the central problem of a visual system is to reconstruct the three dimensional world from the information specified in the two dimensional image on the retina, and, instead, proposed that vision emerges through an organism’s active movement through a visually rich environment. Nativism, strong nativism in particular, claims that certain skills and abilities are innate or hardwired in the brain at birth, and processes external to the individual just play a secondary causal role. For instance, according to nativists, stepping behaviour in infants is determined by a hardwired genetic code, however, anti-nativists challenged this and labelled this idea anti-developmental. That is, according to (strong) nativism, body and environment (including culture) just serve as the “triggers” for ontogenetically determined features that develop in predictable ways. The dynamical systems theory suggests that systems can generate novel behaviours through bodily activity. Its application to developmental psychology offered a radical challenge to traditional nativist assumptions by viewing development as an emergent self-organising product of many decentralised and local interactions taking place in real time. Finally, phenomenology aimed to specify the mechanisms that explain the grounded nature of cognition and the bodily constraints of cognitive agency.

One of the most influential theories of simulation is Barsalou’s theory of perceptual symbol systems (PSS), which is an attempt to implement traditional symbolic functions, such as type-token binding and propositions, using simulation and dynamic systems (Barsalou, 2008; Wilson & Foglia, 2011). According to Barsalou, the problem with traditional theories is that they bear arbitrary relations to their referents in the world. In other words, Barsalou’s argument is that cognitive and perceptual mechanisms essentially involve perceptual simulations which result from the reactivation of multimodal states previously recorded under similar activation patterns. For example, the neural system that represents the colour of an object, originally
formed when the colour is actually perceived, is activated during its absence, such as in imagery or problem solving tasks. A similar theory to that of PSS is Rubin’s basic systems theory (BST) which explains complex memory phenomena such as autobiographical memory (Barsalou, 2008). According to BST, a complex memory is a collection of many multimodal components, including vision, action, affect, language, etc. and memory retrieval involves simulating all its multimodal components together. A further note about simulations is that when they become active they interfere with perceived stimuli. For example, a viewer’s perception of the visual trajectory of an object can go beyond its actual physical motion as a result of simulation interferences. An important implication of Barsalou’s PSS theory, as well as Rubin’s BST theory, is that perceptual symbols or representational states are not independent of the biological system: they commit to the body functions as a constrainer thesis.

Findings related to cognitive linguistics have also shown that understanding the meaning of sentences requires extensive knowledge about bodily capabilities and the possibilities offered by objects, that is, affordances (Barsalou, 2008; Wilson & Foglia, 2011). For example, although the sentence “After wading barefoot in the lake, Erik used his glasses to dry his feet” is grammatically correct, it does not make a sense because the affordances of glasses do not mesh with the action of drying. More examples show that other elements of a natural language, including syntax and semantics, and reasoning in experience, are also grounded in bodily states. Further evidence demonstrated the involvement of simulation (for understanding spatial properties) in comprehending text meaning. For example, individual abilities to comprehend events visually versus verbally are highly correlated (Gernsbacher, Varner & Faust, 1990). Paivio’s dual code theory is an attempt to explain the relationship between language and simulation, and Barsalou (2008) offered a revision by placing deep conceptual processing in the simulation system. According to Barsalou (2008), the powerful symbolic capabilities of humans, compared to nonhumans that lack linguistic systems, emerge from interactions between language and simulation.

The traditionally understood view is that perception is the input from the world to the mind, action is the output from the mind to the world, and thought is the mediating process (called the input-output picture model), however, recent findings show that this is not the case, and have proposed that there is a close connection between perception and action (Barsalou, 2008; Noë, 2004; Wilson, 2002; Wilson & Foglia, 2011). For example, when people were asked to choose from stairs of different heights the one they could ascend most easily, their responses have been consistent with respect to their stair-climbing abilities (Warren, 1894). Similar results have been reported for judgments when grasping objects (e.g. grasping response is slower or faster depending on the orientation of a teapot), catching balls, and climbing walls. As noted above, Gibson (1979) was championed as the first
researcher to reveal this tight perceptual attunement between animal and environment. In support of this view, researchers discovered two visual pathways in the brain, the ventral and dorsal, devoted to “what” and “how” perception, respectively. As a recent continuation of this line of research, Noë (2004) presented his enactive approach to perception in his book titled “Perception in Action.” According to Noë, perception is a kind of skilful (and intrinsically thoughtful) bodily activity on the part of an animal as a whole, but not just a process in the brain whereby the perceptual system constructs an internal representation of the world. It depends on the perceiver’s implicit understanding (a sort of sensorimotor knowledge) of the effects of movement on sensory stimulation. The perceiver enacts (acts it out) the perceptual experience. In support of his claim, he explained the stages in the recovery of normal visual perception after removing a cataract, which does not occur immediately after the surgery, even though the surgery restores the normal visual sensation of the retina (this phenomenon is called ‘experiential blindness’). Noë (2004) rejected the implication, that someone can hypothesise by referring to the function of the dorsal pathway, that perception is for acting or for guiding action, and, instead, demonstrated that the perceiver’s ability to perceive is constituted (in part) by sensorimotor knowledge. For instance, he claimed that an enactive view does not imply that paralysis is a form of (experiential) blindness unless it is more thoroughgoing (such as no eye movements). He argued that mere feeling is not sufficient for perceptual experience, for instance a person who possess only tactile sensation, but has lost all their proprioception and is deaf and blind (i.e. radically inert), will not be able to learn about or discover the properties of objects around them, except, perhaps, simple tactile qualities such as temperature.

Another line of research more closely linked to the embodied approach to visual consciousness is studying how perceived knowledge about tasks or another agent performing a task affects the judgments and performance of an agent (Barsalou, 2008; Wilson & Foglia, 2011). For example, being tired from a run makes a hill look steeper. Simulation theorists, especially social neuroscientists, are increasingly interested in this line of research after the discovery of the brain’s mirror neurons, which are cells with sensorimotor properties that fire both when performing an action and when observing the very same action executed by other individuals. These mirror neuron circuits help the perceiver to infer the goals and affective states of others, such as pain and disgust, as well as a variety of other social activities including imitation and social coordination, however, individual differences in simulation ability (e.g. implicit sensorimotor knowledge) produce differences in social cognition. For example, individual differences in expertise, such as in ballet, correlate with the ability to mirror relevant action (Calvo-Merino, Glaser, Grèzes, Passingham & Haggard, 2005).

While the above theories and empirical evidence particularly fits with the body as a constraint on cognition thesis, it can be shown that the body facili-
tates the regulation of cognitive activities over time and space (Wilson, 2002; Wilson & Foglia, 2011). An early study in line with this thesis is the work of Brooks (1991) who presented a framework to build robots that can respond to real-time demands from the environment, a necessary requirement of any ecological agent, where traditional artificial intelligence has failed with its symbolic framework. Researchers use the phrase “representational bottleneck” to refer to the imbalance between the lack of time to build a full-blown mental model of the environment from which to derive a plan of action, and situations that demand fast and continuously evolving responses. According to Wilson (2002), the human cognitive system has evolved to cope with the representational bottleneck using an effective technique: given an opportunity, humans switch into off-line cognition to reflect upon and plan their further actions (and otherwise, just depend on implicit sensorimotor knowledge). Gesturing while speaking, can also be identified as an evolutionary technique for coping with real-time constraints during communication and language processing. As Wilson explains, gesturing helps to “grease the wheels of the thought process that the speaker is trying to express” (Wilson, 2002, p. 629).

The third role the body plays in cognition, the body as distributor for cognitive processing, reflects that cognitive activity is not limited to the mind alone, but is distributed across neural and non-neural structures: across the entire interacting situation including mind, body, and the environment (Wilson, 2002; Wilson & Foglia, 2011). For example, the shape of bat ears reflects the computational role of echolocation. Further evidence supporting the thesis of body as distributor comes from non-nativist research on perception and learning. In a new-born, the cortical pathways and sensory modalities are domain specific and highly specialised for most tasks, rather than being differentiated as are those of mature brains, for processing particular stimuli. This maturation occurs as a result of an agent’s bodily interactions with the environment.

2.5.2 Human emotion

Until recently, the study of emotion was largely ignored in the mainstream of cognitive science, which considered it a side issue and the antithesis to rational thinking, however, in the last 10-20 years, there has been a rapid increase in studies related to emotion, as evidence emerges supporting the close relationship between cognition and emotion, such as the role of emotions in learning (Picard et al., 2004) and the role of emotions in memory (Galotti, 2004, p. 182). According to Phelps (in Ziemke & Lowe, 2009), there are five types of interactions between emotion and cognition: emotional learning, emotion and memory, emotional influences on perception and attention, emotion in processing social stimuli, and emotion in processing higher cognitive functions.
Although there is a general agreement about what constitutes cognition (i.e. perception, attention, memory, language, etc.), the same does not hold for emotion. In the broad sense, ‘emotion’ is commonly used interchangeably with ‘affect’ to include drives, motivations, feelings, and moods, however, some investigators prefer narrower definitions such as “emotions are states elicited by rewards and punishments.” Some approaches focus on the basic emotions (e.g. Ekman’s six basic emotions: anger, disgust, fear, happiness, sadness, and surprise), while others focus on an extended set of emotions. Emotion theorists are interested in two functions of emotion: evaluative or interpretive judgments about a person’s general state (as emphasised in cognitive appraisal theories of emotion) and bodily reactions (i.e. behavioural and expressive reactions such as facial expressions, gestures, and behavioural choices, as well as somatic markers and physiological reactions such as changes in heart rate and blood pressure) (Eysenck & Keane, 2005, p. 571; Hudlicka, 2009; Thagard, 2005; Ziemke & Lowe, 2009).

While the above description provides somewhat distinct functions of emotion in contrast to cognition, according to most recent work, cognition and emotion are inseparable and very closely intertwined at all levels (Eysenck & Keane, 2005, p. 571; Ziemke & Lowe, 2009). For instance, according to Ziemke and Lowe (2009, p. 114), the embodied view of cognition and emotion suggests that it is “grounded in multiple levels of affective and homeostatic bodily regulation, including motivations, drives, metabolic regulation, etc.” There are still certain open research issues, however, such as the role of emotion in social interaction, individual differences in emotions, the connection between emotions and consciousness, and the physiology of emotion, especially understanding neuronal connections between cognition and emotion (Ziemke & Lowe, 2009).

2.6 Conclusion

This chapter is basically grounded in the high-level theoretical foundation of the research problem addressed by this thesis (while the published papers I-VI are meant to present more specific literature). Fulfilling that intention, this chapter has first explored the domain to which this research belongs: simulation-based training, identifying similar work and elaborating on research issues in that domain. As a result, it identified the critical role played by the psychological fidelity of simulations in increasing the engagement and effectiveness of simulation-based training, however, as the review has failed to identify a theoretical explanation for describing connections between such elements, training effectiveness, and methods that can reliably and effectively capture psychological elements, the review has been expanded towards several other interacting domains.
This chapter has identified three primary domains as the interacting domains: educational psychology, cognitive neuroscience, and computation. The review of educational psychology has helped to identify elements such as the limitations of existing theories of learning for explaining individual experiences in simulators, the connection between those experiences and the learning of skills, and the difference between the learning context and the operational context in terms of transfer of training. The review has helped to recognise that the existing assessment techniques of learning have limitations when applied in simulation-based training contexts, such as in the measures needed and how those measures should be integrated. Expanding the literature review to several other theoretical domains in educational psychology such as emotions in learning has helped to identify several key elements that can be used to understand individual training experiences in simulation-based contexts, however, the explanations found in educational psychology seemed to be very complex and scattered into several theoretical positions without many connections between each other. There was thus a need to expand the review, focusing on theories which can give a fundamental understanding of training phenomena in simulation-based contexts. Before moving into those theories, the chapter explored methods driven by technology to capture elements of the mental state of learners, such as EEG and GSR. The literature review of educational psychology identified the necessity of incorporating information about the mental states of trainees when making a proper assessment of training competence, as those mental states stand as the causal precedents for behaviours. For example, for directing attention during a crash avoidance situation, a driver needs to be in an appropriate mental state, in addition to possessing adequate motor skills.

Expanding the literature review focusing on explanations found in cognitive science has helped to identify two types/levels of explanations attributed to traditional and contemporary cognitive scientific perspectives. Although traditional cognitive science is still in the mainstream, it has major weaknesses with respect to the research problem of this thesis, including poor explanations about the connection between cognition, emotion, and action and between abstract mental states and biological states. Those explanations are still important, however, as most related work in simulation-based training has been conducted based on traditional views of cognition. For instance, evaluating the effectiveness of transfer of training in terms of performance or subjective measures alone is based on the assumptions that the mind plays the central role in cognition, there is no contextual effect for cognition, and there is no interaction between mental states and bodily states. Those explanations are able to explain low level skills, such as memorising road signs and control of the manoeuvres of drivers, however, they face limitations when they are to explain complex phenomena such as the behaviour of an expert driver during a crash avoidance situation, due to the uniqueness and the breadth of mental factors interacting in such situations. On the other
hand, contemporary theories are superior in that they particularly try to ex-
plain the real life behaviours of humans, using psychological, behavioural,
and biological grounds. Those theories not only provide justifications, but
also new directions for incorporating different measures of the mental states
of learners with traditional measures, for more effective and holistic under-
standings of different learning behaviours.
3 Methodology

3.1 Introduction

This chapter presents an overview of the methodology of this research, including its philosophical framework (i.e. ontological and epistemological beliefs) and the low level plans and procedural guidelines of the research process, however, an interested reader should refer to the published papers for more specific details about the methodological elements of each piece of research work.

3.2 The ontological and epistemological stance

Science allows the freedom for a researcher to select the most suitable scientific approach for answering the research question(s) of a study, after making a proper justification of the researcher’s ontological and epistemological stance (Blaikie, 2007; Guba, 1990, p. 17; Kerlinger & Lee, 2000, p. 8; Walsham, 2002, p. 104). Ontology is concerned with the nature of what exists, or the nature of ‘reality’ (Blaikie, 2007, p. 13; Guba, 1990), whereas, epistemology is concerned with knowledge and justified beliefs in terms of their creation and dissemination (Steup, 2014). Since this research is also based on such a belief system, the following paragraphs summarise the ontological and epistemological stance of the researcher in different elements of this research.

From the ontological standpoint, the researcher identified three ontological positions that coincided with this research: 1) the ontological position needed for describing trainees and their interacting training contexts, 2) the ontological position needed for describing the behaviour of performers and their operational contexts, and 3) the ontological position that is left for a researcher who makes interpretations through measures of certain attributes of these individuals and their interacting contexts. The researcher’s belief is that these three types of situations can be best explained using relativist, realist, and critical realist ontologies, respectively. A relativist ontological view is required for describing the learning phenomena of trainees considering the individualistic or subjective nature of learning, and the simulator
experience and the critical role played by the experience in learning (see Section 2.3.1), however, once a trainee becomes proficient or expert, they have to perform in the real world, which is a shared context, and should respect certain common interests. This research associates a realist ontology for describing such performers and their operational contexts. Finally, a researcher who studies the transition of such an individualistic phenomena to a social/shared phenomena has to deal with many challenges, including the subjectivity/objectivity and the interpretation complexity of those observations/attributes. Since this requires the acceptance of the ‘real world’ but at the same time the difficulty faced by the researcher to deal with realism, this research attributes this phenomena to a critical realist ontology.

This research next identified two epistemological positions required for describing, 1) the knowledge of trainees/performers, and 2) the scientific knowledge constructed by the researcher. According to Truncellito (2014), knowledge exists in human minds as mental states known as beliefs; however, only those beliefs which are true and arrived at in the right way constitute knowledge. Although the truthfulness of knowledge is an ontological concern, justifications can be constructed using two approaches called internalism and externalism (ibid.). Internalism assumes that a justification depends solely on factors internal to the believer’s mind, such as someone’s beliefs about the world, sensory inputs, and beliefs about the relationships between their various beliefs. Externalism maintains that at least some factors which have governed those justifications are external to the believer’s mind, such as the reliability of sources. This research has identified the importance of both these processes for the construction of justified and reliable knowledge, however, the researcher rejects the traditional conception of explicit knowledge, at least when describing the knowledge of trainees/performers (see Section 2.5.1 for the role of implicit sensorimotor knowledge in cognition).

3.3 A critical reflection on the research process

The previous section presented the challenge faced by the researcher when deciding the ontological and epistemological stance of this research. Section 3.4 presents several other challenges encountered during the research process. As a result, this research ended up with several iterations in the process of refining the research strategies and conceptual framework as a response to new observations and circumstances (Rudestam and Newton (2007) have indicated that research is not a closed system; Shoemaker, Tankard & Lasorsa (2004) have indicated that the spirit of scientific research is continual self-questioning). In other words, during the research process, the researcher went through different types of reasoning, including induction and deduction (Rudestam & Newton, 2007). In short, induction is the pro-
cess of drawing from specific observations to broader generalisations or theories, and deduction is the process of drawing from general statements (or hypotheses) to specific conclusions. The interaction between the two types of reasoning in a typical dissertation process has been explained by Rudestam and Newton (2007) using the research ‘wheel’ as depicted in Figure 7.

![Diagram of research wheel](Image)

*Figure 7. The research wheel (adapted from Rudestam & Newton, 2007, p. 5)*

As a result of the research process, apart from fulfilling the research problems of this study, there have been certain changes in the researcher’s (enquirer’s) psychology (see constructivist epistemology). As a result, the researcher became more and more sensitive not only to the research instrumentation, but also to the (qualitative) observations he made during the experiments. For instance, during Study III (Section 4.4) he noted better driving performance in novice drivers than in driving instructors, especially in racing on tracks in the simulator; a higher frequency of simulator (cyber) sickness in experts than in novice drivers; and one exceptional complaint by a novice driver that the force feedback he experienced in the simulator was different from that in his previous driving session which had been conducted a few days before (the complaint was indeed true). Although the researcher neither expected nor planned to collect data on those observations, he used those insights to enhance the conceptual framework of this research (according to Creswell (2013), this approach is called ‘explanatory sequential mixed methods’).

In addition to addressing research related issues, the researcher also maintained a consistent record of issues of validity and trustworthiness. Table 3 summarises the evaluation criteria commonly used in both quantitative (positivist) and qualitative (constructivist) research (see Bloomberg and Volpe (2008, p. 76) and Sugarman (2004, p. 45)).
Table 3. General criteria for evaluation for quantitative and qualitative research

<table>
<thead>
<tr>
<th>Criteria for validity in quantitative research</th>
<th>Criteria for trustworthiness in qualitative research</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Internal validity:</strong></td>
<td><strong>Credibility:</strong></td>
</tr>
<tr>
<td>To ensure that the variations in a</td>
<td>To ensure that the respondents and the</td>
</tr>
<tr>
<td>dependent variable are only affected by the</td>
<td>researchers are seeing the world from a shared</td>
</tr>
<tr>
<td>controlled variations in an independent variable.</td>
<td>standpoint.</td>
</tr>
<tr>
<td><strong>External validity (generalisability):</strong></td>
<td><strong>Transferability:</strong></td>
</tr>
<tr>
<td>To what extent the research findings are</td>
<td>The fit or match between the research context and</td>
</tr>
<tr>
<td>generalisable to other groups and settings.</td>
<td>other contexts as judged by the readers, derived</td>
</tr>
<tr>
<td><strong>Reliability:</strong></td>
<td>from “thick descriptions” by the researcher.</td>
</tr>
<tr>
<td>Deals with the stability and replicability of</td>
<td><strong>Dependability:</strong></td>
</tr>
<tr>
<td>research findings.</td>
<td>Deals with the ability to track processes and</td>
</tr>
<tr>
<td><strong>Objectivity:</strong></td>
<td>procedures used to collect and interpret the</td>
</tr>
<tr>
<td>Deals with the naturality of the research:</td>
<td>data.</td>
</tr>
<tr>
<td>the separation between the researcher and the</td>
<td><strong>Confirmability:</strong></td>
</tr>
<tr>
<td>collected data (without manipulations).</td>
<td>Same as objectivity: ensuring that the data,</td>
</tr>
<tr>
<td></td>
<td>interpretations, and findings are rooted in</td>
</tr>
<tr>
<td></td>
<td>contexts and persons other than the researcher.</td>
</tr>
</tbody>
</table>

Since this research incorporated a mixed methods approach (see Section 1.3), it has satisfied the above (Table 3) evaluation criteria, that is, by triangulating different theoretical approaches (educational psychology, embodied cognition, etc.), data collection methods (both quantitative and qualitative data), and standard data analysis methods to increase the credibility/validity of the research. Next, this research addressed generalisability through the use of statistical methods and proper justifications of its research procedure. For instance, when it was necessary to increase the group size of expert drivers in Study III (Section 4.4) by combining both driving instructors and licensed drivers, this was done after a proper evaluation of their driving experiences and other factors deemed to affect their expertise as inferred from the literature. This research has provided a breadth of description about the research contexts and factors under investigation as a way to satisfy the generalisability of the findings. The requirement for reliability/dependability in this research was addressed in various ways. For instance, although the literature has suggested various methods with which to analyse EEG data, the researcher realised that most of those methods were unstable, and, as a result, has incorporated an improved and reliable method (Sections 4.4). This thesis has provided detailed descriptions about the exact methods of data collec-
tion, analysis, and interpretation to satisfy the reliability/dependability requirement. Finally, the requirement for objectivity/confirmability in the research has been fulfilled through detailed and traceable explanations about the research process and the methods incorporated.

3.4 Reflections on major challenges

This section presents a reflection on the major challenges and how the researcher dealt with them during the research process.

Challenge 1

_A theoretical dilemma about the relationship between student engagement and the productivity of learning_

According to certain theories in educational psychology, such as the “attention curve” (Gibbs & Habeshaw, 1992, p. 26), student engagement in a learning task has a direct relationship to the productivity of learning. Several other educators have emphasised the need for appropriate student engagement in a learning task for achieving a more productive learning outcome (Biggs & Tang, 2007, p. 31; Carini et al., 2006). Certain literature in cognitive psychology also supports the above claim (see Galotti (2004) about specialised memories). During the initial stage of the study, the researcher developed a taxonomy for capturing user engagement based on psychophysiological measures (published in Paper I), however, later, as a result of continuing to read the literature, the researcher identified certain factors that put constraints on a student’s ability to concentrate on a learning task such as motivation, interest, mood, and task challenge (Picard et al., 2004). The engagement can be related or unrelated to the expected learning outcomes. For instance, a learner player who plays a serious game can be easily motivated by the entertainment elements of the game while deviating from its expected learning outcomes (see Backlund, Engstrom, Johannesson & Lebram (2008) and Ritterfeld et al. (2009) for an elaborated discussion about serious games). As a result, this research has extended its investigation to: 1) quantify the physical actions of trainees whose activation patterns are determined by the learning situations in the task, and 2) associate performance measures with psychophysiological measures to identify hidden elements of engagement.
Challenge 2

*Identifying a theoretical explanation for simulation-based training*

As discussed in Section 2.2 and elaborated in Section 2.3, simulation-based training differs from traditional classroom-based learning in many ways. Although the literature study has helped to identify several pedagogical elements related to simulation-based training, such as self-efficacy beliefs and flow experience, the researcher failed to identify a unified theory that can explain and connect all those related concepts in simulation-based training. The researcher thus extended the literature review towards certain fundamental theories in cognitive science, seeking explanations, however, during this attempt, the researcher eliminated certain traditional explanations due to their weak explanatory power. For instance, the researcher selected the explanations of embodied cognition of skills acquisition over traditional cognitive science (see Crookall & Thorngate (2008) for a distinction between knowledge and action).

Challenge 3

*Deciding on appropriate sensing equipment for capturing psychophysiological data*

The preliminary experiments of this research (see Section 4.2) were based on collecting data while the subjects were seated in fixed positions, however, as the experiments targeted training environments (see Study III), certain challenges emerged, such as bodily movements and cockpit motion. As a result, certain sensing equipment used in Study I (e.g. GSR sensor) seemed ineffective. There was thus a need to use unobtrusive sensing equipment that could mediate such environmental effects, while being simultaneously capable of detecting a broad range of psychophysiological elements (see Hudlicka, 2009). A further exploration of this issue resulted in the researcher selecting EEG, detected from the scalp of the head, as the most appropriate method (see Section 2.4.1) as the head is the only part of the body with minimum interference from movement-induced artefacts. Next, the researcher investigated potential (e.g. low cost) EEG sensing equipment, including Emotiv Epoc⁴, NeuroSky MindSet⁵, and OCZ NIA⁶ headsets. Since there were no reliable reviews comparing those sensing equipment at that time, the researcher made an evaluation himself (see Appendix B.2). As a result, the researcher identified the Emotiv Epoc headset as the most appropriate (after

---

⁴ [http://emotiv.com/](http://emotiv.com/)
considering a range of factors including the number of electrodes, wireless capability, and access to raw EEG data).

**Challenge 4**

*Uncertainty concerning the validity of Emotiv EPOC neurofeedback headset*

Although the researcher identified the Emotiv EPOC headset as the most appropriate equipment for collecting EEG data, he was confronted by another challenge with regards to the ability of the EPOC headset to pick up ‘true’ or ‘real’ EEG data. He encountered this uncertainty after observing a higher signal-to-noise ratio and failing to find a significant correlation between psychological/behavioural activities and corresponding changes in EEG features, as explained in literature. At that time, certain experts had described the EPOC headset as a ‘toy’, not recommending it for research purposes, however, the researcher was not satisfied with such opinions as there was no strong evidence (e.g. experimental results) as to how they have arrived at such conclusions. As a result, he carried out several experiments and found that the Emotiv EPOC headset can actually pick up P300 responses which is a kind of benchmark for verifying EEG equipment (see Appendix B.2), however, he recognised that the abilities of the headset were still limited due to reasons including lower temporal and spatial resolution, higher signal-to-noise ratio, and inability to apply sophisticated analysis techniques (e.g. independent component analysis).

**Challenge 5**

*Understanding the complex relationships between performance and psychophysiological measures during training*

This research had an ambition to investigate patterns of psychophysiological measures for identifying patterns of training performance. As a result, the researcher investigated the data on the basis of the approach/avoidance performance motivation of drivers (see Section 2.3.3 for theory). For instance, a driver would avoid unfavourable outcomes such as crashing into other cars or walls while approaching favourable outcomes such as higher speed. Although the above investigation helped to identify a formula-driven approach for assessing performance competence in simulation-based training (see Paper III), the researcher could not find a connection between performance and psychophysiological measures during training, however, later, he was able to resolve the issue to a satisfactory degree after using an alternative approach, as discussed in Study III.
Challenge 6

*Understanding the relationship between the features of different psychophysiological data sources*

Challenge 1 described the requirement for performing a proper evaluation of learner engagement in learning tasks, and so it was necessary to extend the work from a simple level of evaluation (Paper I) to a more sophisticated evaluation of engagement. As discussed under Challenge 3, certain psychophysiological measures (e.g. GSR) seemed to be impractical for use in simulation-based training environments. Although the researcher finally decided to continue with EEG, he confronted several challenges as discussed under Challenges 3 & 4. The researcher failed to identify a simple and linear relationship between physiological arousal and the features of EEG as with GSR. Nonetheless, he was interested in such a relationship between the features of brain-based EEG and body-based GSR as it could help to enhance the results of this research. With this intention, he collected both EEG and GSR data during the experiment of Paper IV. According to the literature (e.g. D'Hondt et al., 2010), there are two proposals about the relationship between body-based physiological variations and brain responses. One proposal suggests that emotional stimuli first induces peripheral physiological variations, which occur without consciousness of affect, and the other proposal suggests that the perception of emotional stimuli evokes brain responses that simultaneously but separately induce bodily responses on the one hand, and subjective feelings on the other. Although the data were analysed focusing on a cause-effect relationship between EEG and GSR features, no generalisable and valid result was found, however, the researcher succeeded in using a computational technique based on the second proposal as described in Appendix B.3 (this does not refute the first proposal, rather it can be seen as a limitation of the instruments in attempting to capture such a subtle time duration).

Challenge 7

*Designing an appropriate experiment to evaluate the transfer of training*

The ‘ideal’ experiment for checking the training potential of simulations is to have two groups of trainees (one group, the control group, is trained using a traditional method whereas the other group, the experimental group, is trained using a simulation), and a comparison of performance between those two groups in the real world (this is called 'forward transfer study' or 'predictive validation study', see Liu et al. (2009, p. 55)), however, this type of experiment has certain drawbacks such as being expensive, time-consuming, and risky (especially in domains like aviation and driving, when
the experimental group is tested in the real world). The researcher had the idea that even if such an experiment had been carried out, it would not be possible to conduct a reasonable evaluation, as the simulation-based training domain still lacks effective taxonomies for evaluation. For instance, although an evaluator may conclude that simulations are effective training environments just from observing increased performance in certain overt behaviours, such as lane changing behaviour, in the real world after training in a simulator, the researcher’s opinion is that such decisions are too early. This is because the trainees have been evaluated against a discrete number of overt behaviours, and, even for a given behaviour, the performance has been recorded without considering the variations in internal or external conditions. For instance, a trainee who reported a higher performance in lane changing in a reasonably low traffic condition could end up with a worse performance when the traffic congestion increases or when their mental conditions become unfavourable. A ‘true’ or a reasonable evaluation should thus focus on enduring and holistic behavioural changes rather than those on which traditional methods focus.

The researcher’s suggestion to address the above challenge was to have an evaluation process involving a breadth of measures comprising trainee perceptions of situations they have faced. Essentially, those measures should involve the neural mechanisms underlying those perceptual processes; trainees’ own judgments and justifications about those situations and a self-assessment of their performance; and their performance as captured based on overt behaviours which are meaningful for the task (Chapter 2 provides a theoretical foundation for this decision). A similar set of data should be gathered involving experts in that specific skill domain when they perform in both the simulator as well as in the real world (see Brennecke (2009) for the importance of a teacher’s perspective in learning contexts). In this type of arrangement, the trainee performance in the real world is considered optional (according to literature, e.g. Liu et al. (2009), this type of arrangement can be identified as a combination of both backward transfer and quasi-experimental studies). Another advantage of the above approach is that it does not require the performance of trainees to be judged by experts (see Section 2.3 for drawbacks of involving experts including their subjective judgments). A further justification of the proposed approach is given below.

The first real world driving session of a novice driver under the supervision of a driving instructor usually involves a cognitive bottleneck with the process of accurately and efficiently grasping and responding to situations in the environment under time pressure and the interference of the instructor, however, once the driver becomes proficient or expert in driving, they do not usually need to give a second thought to this process, rather it happens straightaway and automatically (see Dreyfus (2004) for a discussion about the transition of skills from novice to expert under a phenomenological perspective). The researcher predicted that an individual’s skilful perfor-
mance of an activity on two types of internal processes: consciousness of situations and bodily states, and a proper interaction between them, however, those two processes depend on several other factors as well, such as self-belief, emotions, metabolic conditions, and task-specific motor skills (Section 2.3 provides a theoretical basis for this idea). The researcher believed that the proficiency or expertise of an individual can be evaluated on the basis of the equilibrium between these two processes when that individual engages with the skilled task. In other words, the researcher believed that an individual becomes an expert not only because others say so, or that individual believes they are, but when there is an internal element in that individual which is pervasive throughout their mind and body as a result of honest or hearty reflections on their experiences in the past.

3.5 Experiment design and data collection

As the primary ontological stance of this research was determined as critical realism (Section 3.2), this research preferred experiments under natural settings (i.e. quasi-experiments). This research used a mixed methods approach (Creswell & Clark, 2006; Creswell, 2013, p. 4; Myers, 2009, p. 8) for collecting both quantitative and qualitative data. Section 3.4 discussed the challenges faced by the researcher when designing the experiments and Chapter 4 presents the details of the four experiments conducted in this research, along with the different types of data collected.

3.6 Data analysis

This research used different types of statistical techniques and computational tools for data analysis. The statistical techniques included simple descriptive techniques such as mean, median, variance, standard deviation, and different graphical representations of data, as well as inferential techniques such as hypotheses testing (e.g. t- and F-tests) and clustering/classification of data points into groups (see Kerlinger & Lee, 2000; Ross, 2004). The computational techniques included the use of different algorithms and software tools such as MATLAB (MATLAB, 2003), EEGLAB (Delorme & Makeig, 2004), and Transana (Anderson, 2008), however, this research has avoided an extensive use of machine learning (ML) techniques because the artefacts present in psychophysiological signals (especially in EEG) can lead to serious misinterpretations of data (Chadwick, McMeekin & Tan, 2011). Some authors also criticise the use of ML techniques for inference (Breiman, 2001). Chapter 4 presents the details of data analysis with respect to each study, and the more specific steps in the data analysis can be found within each published paper.
3.7 Ethical considerations

Research ethics contain certain general principles that any research must follow if it involves human participants (Bloomberg & Volpe, 2008, p. 76; Sugarman, 2004, p. 50). Such principles are required to prevent risks (from stress, coercion, deception, or the invasion of privacy) and promote informed consent in research (Sugarman, 2004, p. 51). These principles and procedures are published by professional associations such as the British Psychological Society\(^7\) and the Swedish Centre for Research Ethics & Bioethics\(^8\). The following list summarises the key ethical principles followed throughout this research:

- The research was designed in such a way as to avoid novice drivers risking their lives in real world driving
- Before the experiment, each participant was given a verbal explanation of the purpose and nature of the investigation and procedures such as the data collection methods and the way that data will be treated in the future
- Informed consent was obtained from each participant before the experiment (Appendices C.1 and C.2)
- The participants were allowed to withdraw their participation at any time without giving a reason (note: some participants withdrew their participation in the simulator sessions as a result of this freedom)
- Personally harmful questions were avoided in the questionnaire
- After the experiment, each participant was asked about their experience while taking refreshments. They received a free lunch and/or cinema tickets as compensation for their time spent being involved in the experiment
- The personal information about the participants was decoupled from collected data when that data was stored in the computer and that personal data was excluded from publications

3.8 Conclusion

This chapter explained the methodology of this research including its philosophical assumptions and choice of methods for data collection, analysis, and interpretation. It explained the measures taken in order to improve the validity, trustworthiness and the ethical principles followed during the research process. Chapter 4 presents the integration of these methods in the actual empirical work.

\(^7\) [http://www.bps.org.uk/](http://www.bps.org.uk/)
4 Empirical work

4.1 Introduction

This chapter presents the empirical work of this research by organising the research work presented in the six publications into three themes of studies. Figure 8 summarises the published work attributed to the three studies along with several high-level objectives implied by the research questions.

![Diagram of empirical work]

Figure 8. The relationship between high-level objectives, publications, and the three themes of studies

Studies I & II are based on the theoretical understanding that student engagement in a learning task has a major influence on the productivity of learning. As discussed in Chapter 2, conventional assessment techniques are basically focused on the product rather than the process of learning, and therefore such techniques become inadequate in the assessment of skills in
training. As the process of training involves both external performance changes as well as internal changes, including psychological and physiological processes, a technique which focuses on the assessment of training should incorporate an evaluation of such internal factors in order to make an accurate judgment of someone’s skills. Although these factors can be assessed using subjective measures (e.g. self-evaluation questionnaire), they are attributed many disadvantages, such as that they depend on someone’s ability and willingness to articulate their own mental experience (already discussed in Chapter 2). The researcher thus identified two alternative ways to determine these factors or to understand the interrelationship between internal and external variables. That is, 1) by using psychophysiological measures, and 2) by capturing how performance measures reflect psychological variables. The former has been evaluated under the scope of Study I, and the latter has been evaluated under the scope of Study II. Later, Study III extended the work by interrelating both internal and external measures. Figure 8 shows the evolution of the conceptual framework of the research after integrating insights of embodied cognition. The preceding sections elaborate on each study presenting and discussing the empirical work conducted.

4.2 Study I: Determining types/levels of participant engagement using psychophysiological measures

Chapter 2 has elaborated the close connection between cognition and emotion and the important role played by emotions in learning. Both Chapters 2 & 3 have emphasised the importance of integrating learners’ internal measures, collected while the learners are interacting with the learning tasks, with other external measures as a method to make more reasonable evaluations of their learning experience. The evaluation of a learner’s internal measures is needed to make an assessment of the psychological fidelity of simulations (discussed in Section 2.2). This study was basically focused on developing an empirical and theoretical foundation to determine the level of engagement of participants in different types of multimedia, and to use psychophysiological measures to identify the mental activation related to participant awareness/engagement levels. As a result, Study I answers the research questions RQ1 and RQ2. Study I also has an intention to understand the body’s role in cognition/behaviour based on relationships between different behavioural measures. Study I thus also addresses RQ3.

The aims of Study I were fulfilled by conducting two experimental studies and publishing the results in Paper I and Paper IV. The following is a summary of the work presented in the two papers and a further discussion of the results.
Paper I: Determining the depth of engagement

This paper reported on an experimental study which has been conducted to identify the type/level of engagement in a set of multimedia types based on psychophysiological measures. The conceptual framework of the work presented was based on certain theories in cognitive psychology, such as that eye movements directed to a location in space are preceded by a shift of visual attention to the same location (Handerson, 1990; Hoffman, 1998; Irwin, 1990). In the study, electrooculogram (EOG) signals were used to determine whether a participant was looking at the computer or not, however, as visual focus is not the sole factor that determines the actual engagement with the corresponding event or object (e.g. attention can be different in a highly alerted state than in a bored state), an arousal evaluation was also performed using GSR measures (see Section 2.4.2).

Experiment setup and data analysis

The experiment was conducted involving six participants (five male and one female; age between 25-35 years) in a lab environment at the University of Colombo School of Computing, Sri Lanka. In the experiment, each participant was shown several types of multimedia content such as still pictures, video clips with/without exciting events, and songs without containing visual content, using a regular computer. Each recording was followed by a brief and uninformed recording for capturing data about the off-screen interactions of subjects. Biophysical signals were recorded using low cost homemade EOG and GSR equipment and a combination of open source and customised software (see Paper I for details). Figure 9 shows the setup of the experiment.
The collected data included psychophysiological measures (i.e. GSR and EOG) and video data (i.e. scenes of multimedia content and the facial expressions of the participants). The data was analysed in MATLAB considering the features of GSR (e.g. mean SCL, variance, and event related changes in SCL) and EOG (e.g. mean and standard deviation) interlaced with different scenes and events in the multimedia content, and each subject’s attention changes as observed from their facial expressions (see Paper I for further details).

**Key results of the paper**

Figure 10 shows the EOG waveforms of on-screen (a video clip) and off-screen (a brief uninformed recording after the computer based interaction) situations for the six subjects who participated in the study.
As can be seen in Figure 10, off-screen situations had increased (magnitudes of 1.5 to 3.5 times higher) EOG activity compared to on-screen situations. Next, the degree of engagement was estimated using the GSR activity. For instance, Figure 11 shows the GSR waveforms of subjects who watched a video clip that contained an exciting event at a particular movement in the media timeline (within 25-35 seconds).

Figure 10. EOG signals of each subject during (left) on-screen interaction, and, (right) off-screen interaction
The key findings of the study are listed below.

- Multimedia content containing emotion-evoking events have the potential to increase the degree of engagement of subjects (as observed from their facial expressions). The psychophysiological features associated with increased engagement included higher variances and moderate levels of GSR as well as changes in the GSR waveforms interlaced with emotional events in the media timeline.

- When the subjects concentrated on media content that contained no or fewer emotion evoking events (e.g. a video lecture), the GSR waveforms became smoother (with respect to high frequency variations) and GSR levels approached peak levels, however, as could be observed from their physical behaviour, including facial expressions, these types of interactions can easily induce inattention or bored states in subjects. When the attention of subjects shifted away from the screen, the means and standard deviations of EOG waveforms increased. GSR levels became lower during off-screen interactions, suggesting an increased level of engagement.

**Paper IV: Comparing two types of scenarios**

This paper reported on an experimental study in which the aim was to investigate whether scenarios based on animated characters are equally capable of triggering psychophysiological activity as scenarios based on real actors. The work was motivated by developing a low cost alternative for the treatment/rehabilitation of mentally disordered offenders (MDOs), which is currently being conducted using verbal descriptions (see Murphy (2006), Rogers, Dziobek, Hassenstab, Wolf and Convit (2007) for limitations of verbal descriptions) or involving expensive video-based systems (e.g. Wijk
et al., 2009). The results of Paper IV are equally important for simulation research as it provided an opportunity to evaluate the virtual scenarios on which most simulations are based. The following is a summary of work presented in the paper.

**Experiment setup and data analysis**

This experiment was conducted in a quiet lab environment at the DSV of Stockholm University, involving 26 subjects (mean = 23.9, SD = 5.3), however, the datasets of eight subjects were abandoned due to errors in data collection. In the experiment, one group (n=11, mean age = 22.2 years, SD = 2.4) watched a film of real actors, and another group (n=7, mean age = 25.1 years, SD = 8.0) watched an animated film, which had the same story and dialogue as the real actor film.

The experiment involved two films, one based on real actors and one on animated characters, both having the same story and using the identical sound track. The beginning of the story consisted of a neutral family situation between a man and a woman which later developed into a tense and violent situation as a result of the man misinterpreting the woman. In the story, the man is supposed to be an MDO. Figure 12 shows screenshots of the two films, representing a certain situation in the story.

![Figure 12. Screenshots of two types of films](image)

The recorded data included GSR and EMG using the BioPac MP150 equipment and AcqKnowledge data logger, EEG data using the Emotiv headset and TestBench data logger, and video data about facial expressions using the Camtasia recorder. Figure 13 shows the setup of the experiment.
The data was analysed in Matlab, considering both phasic and tonic features of GSR: the pulse location (valley), skin conductance level (SCL) at that location, rise time, amplitude (at peak), and recovery time of SCR pulses (see Paper IV for more details).

Key results of the paper
A summary of key findings from the study is given below:

- Both ANOVA and t-test showed no significant difference between the mean SCR scores of the two groups (see Figure 14), suggesting that both types of scenarios are equally capable of triggering psychophysiological activity of subjects.
- Both ANOVA and t-test showed that the mean SCR amplitudes of the animated film group were significantly lower than that for the real actor film group (see Figure 14). This suggested that animated scenarios are
weaker than scenarios based on real actors with respect to the intensity of the evoked psychophysiological activity.

- The Pearson’s correlation coefficient between the two groups for the overall film length of both SCR scores and amplitudes showed that although there was no correlation for overall film length, there were correlations for certain time intervals of the film (see Figure 15). A further analysis targeting those time intervals revealed that they actually correspond to the most critical parts of the story where a user is expected to elicit very strong emotional reactions.

**Figure 14.** Box plots of ANOVA comparing real actor film group [1] and animated character film group [2] on the basis of (left) SCR scores, and (right) SCR amplitudes

**Figure 15.** Correlation coefficients between the two groups SCR scores (solid line) and SCR amplitudes (dashed line) using a moving window of 60 seconds time frame
Discussion of results

The results of Paper I revealed a close relationship between psychological processes of engagement (i.e. attention and arousal) and physiological events as reflected in biophysical signals (i.e. EOG and GSR), that is, the ability to recognise the type/level of a participant’s engagement based on psychophysiological measures. This approach is more reliable and advantageous than most existing approaches, including the methods based on facial expressions and changes in voice (Healey & Picard, 1998; Song, Park & Jeong, 2006). The research provided insights into conducting similar research, especially the methodological elements such as the selection of appropriate equipment and data analysis techniques, environmental conditions which can affect the data, and ethical issues, however, the research had several limitations. For instance, the predicted engagement types/levels have not been verified using a relatively reliable measure such as a self-evaluation questionnaire, the study used less accurate equipment, a low number of subjects, few types of multimedia, and a limited number of features of both EOG and GSR signals.

Paper IV addressed most of the weaknesses that were present in Paper I. For instance, it used more reliable equipment for data collection, set up a clear objective for research (i.e. only to discriminate between two types of scenarios), and increased the number of subjects. Although the results of Paper IV led to the general implication that the depth of engagement is higher in real actor based scenarios than in animated character based scenarios, the researcher’s view is that the results should be interpreted considering certain limitations of the research. For instance, the sound track may have influenced the results because the film based on the animated characters used the sound track which was part of the film based on real actors. There was therefore a greater chance that a mismatch took place between the visual and auditory contents of the animated character based film. The quality of animations was also not of a high standard in the film. These problems may have resulted in the decreased engagement of respondents (see the discussion in Section 2.5.1 about the body as a constraint on cognition thesis). To clarify this dilemma the researcher conducted a limited study which is presented in Appendix B.3. The results of the study showed two patterns of brain activation in those two groups, with the indication that the animated group had experienced a certain difficulty in relating the auditory content to the visual content (future research is required to verify this). If this result answers the above dilemma, it also provides supportive evidence of embodied perception (see the discussion about embodied cognition in Section 2.5.1).
4.3 Study II: Determining the skills of trainees

Although Study I focused on evaluating the possibility of integrating psychophysiological measures into simulation-based training research, the actual scope of that study was limited to determining the engagement of users in emotion-provoking events with different types of multimedia content. In other words, Study I could not achieve its ultimate goal as it lacked an effective framework to differentiate the engagement as related and unrelated to training. The researcher thus designed Study II to investigate the possibility of quantifying the behaviour of trainees and to evaluate those behaviours with respect to their psychophysiological bases (see Challenge 5), to enhance the scope of Study I. In other words, Study II was basically focused on the research question RQ4, establishing a relationship between performance and psychological variables for comparing simulator and real world training. For fulfilling the above aim, two conceptual frameworks were proposed and published in Paper II and Paper III. The following is a summary of the work presented in those papers, highlighting the key results.

Paper II: Game Interaction State Graph (GISG)

The GISG is based on the idea that each trainee, during their training, has to maximise their experience by facing different types of driving situations. For instance, a trainee driver should not only learn to drive safely on the road but they have to experiment (or explore) different driving conditions, such as slow driving and fast driving, to experience the advantages and disadvantages (e.g. boredom, stress, and risk of accidents) of such conditions. The GISG is proposed with the idea of visualising different driving behaviours to capture the extent to which an individual has experienced different driving conditions or situations (see Figure 16). The driving conditions or situations have been identified based on certain theories and findings of driving psychology research. For instance, a driver at low speed, driving on an empty motorway in otherwise good conditions, will find it difficult to concentrate on the task while feeling bored (Collet et al., 2005; Lal & Craig, 2001; see also http://www.safespeed.org.uk/). The driving conditions or situations were categorised based on their relative significance for training effectiveness (called the risk factor, which varies from 0 to 5 in Figure 16). Finally, a driver’s behaviour is captured in the graph, based on a set of predefined behavioural rules of which the activation patterns are determined by the trainee’s physical actions and the state of driving scenarios.
Paper III: A formula based on performance competence

This approach is based on the relationship of a trainee’s adaptation of goals used to approach favourable outcomes (F), to avoid unfavourable outcomes (U), and the physical effort (E) the trainee makes to satisfy the above expectations. The relationship was further extended to distinguish a trainee who achieves a certain performance putting in less effort, as compared to another trainee who achieves the same training outcome by putting in greater effort, by introducing the concept of optimal effort (O) which is attributed to an expert in that skill domain. The above relationship has been expressed as a formula to obtain a numerical value of performance competence of a trainee as given below.

\[ P = \frac{(F - U)}{(E' - O)} \]

For validating the proposed formula-driven approach, an experiment was conducted to gather data as shown in Figure 17. This experiment was conducted in a quiet room environment at the University of Skövde. In the experiment, each participant (four males and three females; aged between 25-44 years; mean age of 32 years) played an open source car simulator game with the video output projected onto a large whiteboard, and using physically realistic steering wheel and pedals. Although 14 sessions were recorded, one recording was dropped due to errors in the recording, which resulted in 13 session recordings for the seven subjects. The recorded data included players’ physical actions (throttle, brake and steering wheel angle), driving variables (vehicle speed, crashes and position in the track), EEG data, and video data regarding driving scenes and player positions.

The values for the formula were determined by considering the vehicle speed as a favourable outcome, both the scores of crashes and off-track driv-
ing as unfavourable outcomes, and measures of throttle, brake and steering wheel angle as constituting effort. A value for the optimal effort was determined by trial and error. In addition to the quantitative approach proposed in the paper, a qualitative approach (analysing video data using pre-determined criteria) was used as a means of comparing results. As a result, the correlation analysis (based on Pearson’s correlation) between the results of the proposed formula and the qualitative method, reported both high (e.g. 90.25%) and low (e.g. 13.52%) coefficients for different types of road segments. For instance, Figure 18 shows the values obtained for two types of road segments (straight roads and turns/bends) using the proposed formula and the qualitative method.

Figure 18. Correlation analysis results for different types of road segments.
Discussion of results

The two methods discussed above, the GISG approach and the formula-driven approach, have been proposed to distinguish the behaviours of trainees with respect to their training domain, based on performance measures, however, there are certain differences between the two approaches, including advantages and disadvantages. The GISG is a graph which shows the transition of different behaviours of a trainee during their training process. Based on this graph, decisions can be made such as a trainee’s natural tendency to experience different conditions or situations in a given skill domain. A GISG thus provides an opportunity to alter or guide a trainee to maximise their experience. A GISG can also be used to capture certain outstanding skills of trainees or to compare the skills of a trainee with the skills of an expert based on the patterns of behaviours. The main disadvantage of the GISG based approach is the requirement for behavioural rules and identifying a set of performance measures that accurately describe the activation pattern of those rules. As a further improvement to this approach, psychophysiological measurements can be infused into the behavioural rules for the purpose of accurately capturing the mental conditions that resulted a certain type of behaviour. For instance, such rules should be able to differentiate the behaviourally similar crash avoidance situation of an active novice driver from a frustrated advanced driver because the latter can be more risky than the former, depending on the psychophysiological activity of the driver.

The formula-driven approach gives the numerical value of performance competence of a trainee as a continuous variable which can also be used on sections or sessions of a task by averaging the values. This approach is more robust than the GISG-based approach, as it is grounded in more formal theories of motivation, however, there is some speculation about the validity of this method as the results have reported both high and low correlation coefficients between the proposed approach and the qualitative method used. The
researcher’s answer to this quandary is twofold: the high correlation coefficients between the two methods already justify the validity of the proposed approach, however, the low correlation coefficients could be associated with an improper combination of performance measures to capture the behaviour of the subjects. The researcher proposes a thorough future evaluation of the proposed method by increasing the number of subjects and types of road segments as well as by carefully selecting the parameters of the formula and using a more reliable (qualitative) method to validate the results.

4.4 Study III: Comparing the behaviour of trainees

Studies I and II aimed to develop a conceptual and methodological foundation for carrying out a more focused investigation into the research problem of this thesis, however, the preliminary attempts at associating psychophysiological measures with performance and subjective measures in learning environments have not resulted in successful findings (see Challenge 5). As a result, the final stage of the investigation pertains to Study III, and was driven by a revised conceptual and methodological framework along with a larger experiment, with two objectives: 1) to compare the driving behaviour of experienced drivers between on road and driving simulator (as presented in Paper V), and, 2) to compare the driving behaviours of novice and expert drivers in the simulator (as presented in Paper VI). The following is a summary of work presented in those papers highlighting the key results.

Subjects

Table 4 summarises the details of the three categories of drivers who participated in different tasks of the experiment. Both driving instructors and novice drivers were recruited from two driving schools, and the regular drivers were recruited from university staff. Each participant signed a written statement of informed consent (see Appendix C.2) prior to their participation in the experiment and received a free lunch, refreshments, and cinema tickets as compensation for their time spent being involved in the experiment.
Table 4. Details of the three categories of drivers who participated in Experiment 4

<table>
<thead>
<tr>
<th>Sessions and tasks of driving</th>
<th>Driver category, age, and number of participants for each session</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Driving instructors</td>
</tr>
<tr>
<td></td>
<td>Five male and three female aged 27 to 56 years (mean = 40.9, median = 43.5, SD = 11.5)</td>
</tr>
<tr>
<td>RW-T11</td>
<td>8</td>
</tr>
<tr>
<td>RW-T21</td>
<td>8</td>
</tr>
<tr>
<td>RW-T22</td>
<td>8</td>
</tr>
<tr>
<td>RW-T13</td>
<td>8</td>
</tr>
<tr>
<td>S1-T30</td>
<td>5</td>
</tr>
<tr>
<td>S1-T41</td>
<td>7</td>
</tr>
<tr>
<td>S1-T51</td>
<td>6</td>
</tr>
<tr>
<td>S2-T30</td>
<td>6</td>
</tr>
<tr>
<td>S2-T41</td>
<td>6</td>
</tr>
<tr>
<td>S2-T51</td>
<td>7</td>
</tr>
<tr>
<td>S3-T30</td>
<td>0</td>
</tr>
<tr>
<td>S3-T41</td>
<td>0</td>
</tr>
<tr>
<td>S3-T51</td>
<td>0</td>
</tr>
</tbody>
</table>

Note: The naming convention used in labelling different driving tasks/sessions are as follows. Real world driving tasks are prefixed with ‘RW’ and the number proceeding ‘T’ (or ‘Tr.’) represents different parts of the driving task. That is, ‘T11’ and ‘T13’ represent onward and return trips of city traffic driving, respectively, and ‘T21’ and ‘T22’ represent onward and return trips of highway driving, respectively. In the simulator driving tasks, the number proceeding ‘S’ represents the session and the number proceeding ‘T’ represents each of the three types of tracks. That is, ‘T30’ represents the highway traffic track, ‘T41’ represents the Monaco track (a track in a city area without traffic), and ‘T51’ represents the LeMans track (a racing track without traffic).

**Experiment setup**

This experiment was conducted involving both the mid-range driving simulator at the University of Skövde and real world driving. Figure 19 shows the setup of the experiment.
The driving scenarios are projected on a screen in the case of driving simulator.

Sending information about driving scenes

Capturing vehicle’s physical performance data through sensors and ATMega16/32 microcontrollers

Recording vehicle’s physical performance data, driving scene data, and video data in a synchronized manner

A custom software for logging vehicle’s performance and EEG data in a synchronized manner

Figure 19. Setup of Experiment 4
Note: The car used for real world driving was equipped to collect the same data as did the driving simulator.

The real world driving task consisted of driving in both city traffic and on a highway for about 20 minutes, and due to the risks and ethical issues only the driving instructors participated in that part of the experiment. In the simulator, all 31 participants drove on several different types of tracks that represented conditions both with and without traffic, as well as both city-like and highway-like roads (see Figure 20). The simulator driving experience consisted of up to three driving sessions on three different days.

Data gathered
The collected data included performance measures (e.g. throttle, gas, and brake pressure, steering angle, vehicle speed, and vehicle position on the road), psychophysiological measures (i.e. EEG and heart rate), and subjective measures (i.e. ratings and answers to questionnaire, and video data).

Data analysis
The analysis of data was carried out by identifying features of different categories of measures which included means and standard deviations of per-
formance measures, time-frequency analysis of EEG and EEG-based vigilance estimators, and subjective ratings of their driving experience in the simulator (see Appendix B.1 for details about data analysis). The analysis involved computational tools such as Matlab and EEGLAB and statistical tests such as ANOVA F-test.

Key results of the study
A summary of the key results and the findings of Study III is given below.

- No significant difference has been identified between driving instructors and regular drivers in personal data, driving experience, and driving performance in the simulator (using ANOVA F-test with p-value < 0.05). This result justified the decision to combine both driving instructors and regular drivers in a single category called ‘expert’ or ‘experienced’ drivers.
- The results showed no significant main effect from different driving sessions in the simulator (using ANOVA F-test with p-value < 0.001), however, there were main effects from different tracks in the simulator and from two types of drivers (i.e. novice and expert).
- The results showed well-mannered and gentle driving behaviour of experts in the real world than in the simulator (based on the spread of points in the scatterplot of Figure 21 and standard deviations of driving variables as presented in Paper V).
- According to both quantitative (i.e. performance and subjective ratings) and qualitative (e.g. getting nausea and personal opinions) results, the first simulator driving experience of the expert drivers was more difficult for them than for the novice drivers.
- The novice drivers performed much better (speed and time to complete a track) than experts on the tracks which were free from other traffic and welcomed speeding (as captured by performance measures such as speed), however, the order was reversed on the highway traffic track (T30) on which the experts performed much better than the novice drivers.
- According to the results of combining both performance and psychophysiological measures, driving on the highway traffic track (T30) was more stressful for novice drivers (about 2 to 4 times higher); and more manageable for expert drivers (considering both VG and TD measures). On the other hand, the experts experienced stress (2 to 3 times higher stress than the novices) while the novice drivers performed within their comfort range on the other two tracks (T41/51).
- The experts exhibited more adaptive behaviour (with respect to translating their existing skills) as they were exposed to the consecutive sessions in the simulator as seen in their performance measures, as well as subjective ratings (e.g. flow), however, the performance orientation was more visible in novice drivers in all three types of measures on tracks.
T41 and T51, except on T30 in which the features (e.g. other traffic) and challenge of the task placed some psychological stress on the subjects.

- The infusion of psychophysiological measures into subjective ratings showed that flow was associated with focused concentration and learning (based on the activity of the gamma band), whereas tension was associated with both flow and the normal state of wakefulness (based on the activity of both gamma and beta bands).

Figure 21. Scatterplot between means of gas (MGS) and means of speed (MSP) for different road types in the real world (i.e., Tr.11, 13, 21, and 22) and in the simulator (i.e. Tr.30, 41, and 51) for experienced/expert drivers
Discussion of results

The results of Study III confirmed that driving behaviour is highly task-dependent and depends on the characteristics and conditions of drivers (e.g. age and experience) and the environment (see Mullen et al. (2011), p. 2 for a similar discussion). For instance, both experienced and novice drivers behaved differently in different driving tasks. These results can be understood both in support of simulation validation and the limitations of the research as discussed below.

This research involved a combination of both backward transfer and quasi-experimental studies for simulation validation after considering the risks of involving novice drivers in on-road driving (see Liu et al., 2009, p. 56), however, the credibility of this approach relies on certain assumptions, including the drivers believing in the authenticity of the simulation (Caird & Horrey, 2011) and the characteristics of drivers, such as there being no significant age difference between the participants. Since this research faced a practical challenge in finding participants in the same age group, it relied on two categories of drivers whose mean ages were significantly different: the mean age of novice drivers was significantly lower than the mean age of expert drivers. Although this is the natural tendency in society with regards
to drivers, this has consequences for the subsequent driving behaviours of the participants. Young drivers are generally characterised by risky or over-stimulated driving behaviour whereas older drivers are more cautious in their driving behaviour (Ouimet et al., 2011). Similar driving behaviour has also been observed in this research when comparing the driving behaviours of novice and expert drivers in certain tasks in the simulator, however, the behaviours were also influenced by belief in the authenticity of the simulation. Although believability depends on the fidelity (especially the psychological fidelity) of simulation to a great extent (see Backlund, Engstrom, Johannesson, Lebram, et al., 2008), it is not possible to believe that a participant is completely unaware that a simulation does not have real world consequences. This explains the well-mannered and gentle driving behaviour of experts in the real world compared to that in the simulator, however, there were several other results that were not in line with the above argument. For instance, the expert drivers proved to perform better than the novice drivers when the driving task had similar features to that of real world driving (e.g. traffic congestion and the terrain). This indicated the true skills of the experts and the simulator’s ability to challenge the trainee drivers if a simulator can offer real world like tasks. In other worlds, a carefully designed training simulation (e.g. having an appropriate psychological stress) has the potential to offer a real world-like training experience.

Apart from the above results, this research did not find a significant effect on driving behaviours of subsequent events in the simulator, other than a change in certain high-level opinions as expressed in the questionnaire (see Caird & Horrey (2011, p. 11) for common threats to internal and external validity). This may be because the events (e.g. traffic) appeared as part of scenarios in a random order, with random features. With respect to the similarity between the behaviours of simulator and on-road driving, this research has captured the fact that, in general, the participants were more aroused (had high vigilance) in the simulator than on-road. This could be interpreted from several perspectives. Most drivers are driving in the real world for a routine purpose (e.g. going to office) and in their own vehicle; however, the simulator is an unfamiliar environment in which most of the events are artificial (which can trigger a higher degree of cognitive disequilibria or curiosity, resulting in an active mind). This results in a fundamental concern about training transfer to on-road skills. For instance, a driver who has performed much better in crash avoidance situations in a simulator may not perform well in on-road situations if their body is not prepared to cope with such situations, however, a driver may be able to minimise the damage as a result of their training to anticipate hazards.

Another view of high vigilance driving behaviour in the simulator is via its dependence on the type of engagement of the participants. It is more likely that the participants have adopted an entertainment oriented performance motivation as part of their engagement in the simulated tasks which is an
expected behaviour in most serious games (see Ritterfeld et al., 2009, p. 5). After considering the above results, it can be concluded that simulators are not adequately able to provide experiences which are of the same standard as in the real world, however, simulators can provide training opportunities for anticipating hazards by determining the limitations of trainees and to improve them (see Caird and Horrey (2011, p. 7) for a similar discussion).
5 Conclusion and future research

5.1 Introduction

This research set out to investigate a research problem in the area of simulation-based training, with a focus on the validity of simulations, especially for skills training such as driving skills. In other words, this research has investigated the user experience of trainees with respect to the acceptance of virtual scenarios for provoking a similar psychophysiological response as in real world scenarios and the training potential of simulations with respect to the positive transfer of training from a simulator to real world operational contexts. The requirement for this research emerged as most existing work has failed to determine which elements of training simulations would contribute to an increased transfer of training, and the weaknesses of existing methodologies to detect a broad range of elements related to human factors in training. By fulfilling the above aim, this thesis has reported on an investigation based on four experimental studies and six publications. As the final part of this thesis, this chapter attempts to explicitly demonstrate that this thesis has indeed accomplished its goals by making a distinct contribution to the body of knowledge, with further implications as elaborated below.

5.2 Conclusions about the research questions

This section discusses how this research has answered the four research questions asked at the beginning of the thesis.

RQ1:

*How can psychophysiological measures be used to determine the mental activation related to a participant’s awareness/engagement level in the simulator?*

This research has answered the above research question to different levels and degrees. At a basic level, Study I examined how psychophysiological measures can be used to determine the depth of engagement to a set of mul-
timedia types, and to compare two types of scenarios based on certain features of psychophysiological measures. The results of Paper I indicated that EOG can reflect a participant’s attention to visual content, whereas GSR can determine the depth of engagement based on how the patterns are interlaced with events in the media. Paper IV reported on a study, improving the methodological framework of Paper I, comparing scenarios based on real actors and those based on animated characters, based on the features (i.e. frequency of the occurrence of SCR and their amplitude) of GSR.

In Study III, psychophysiological measures were used to determine the vigilance level (or sustained attention) of a participant while they are engaged in a driving task. Since the sensing equipment used in Study I was determined as inappropriate for use in Study III (e.g. noise due to bodily movements), it relied on EEG based measures (see Challenge 3). Study III eliminated the difficulties of analysing EEG by using an improved computational technique (discussed in Appendix B.1).

RQ2:

What features of psychophysiological measures can be utilised to generalise about individuals and experimental conditions in simulator validation?

This study answered the above research question depending on the type of measure and the nature of interaction of participants. In Study I, the interaction was limited to observing a few types of media (e.g. short films) while the participants remained seated in front of a computer in a fixed position. This enabled the researcher to use simple measures, including GSR and EOG, as there were very low interferences from bodily motion. Although GSR and EOG have relatively simple and linear relationships with psychological processes (Andreassi, 2013; Dawson et al., 2007), the researcher had to overcome individual differences of psychophysiological measures. In Paper I, this was addressed by observing the tonic changes of GSR waveforms (i.e. SCL changes) interlaced with known events (e.g. an exciting scene) in the media. There was a consistent change in the magnitude of EOG signals over each participant depending on their visual focus, however, as there were no annotated events in the films used in Paper IV, it had to rely on a different method. Instead of analysing tonic changes, Paper IV analysed phasic changes of GSR waveforms (i.e. SCR pulses) identifying when they occur and their magnitude (amplitude). The results were obtained by averaging the two types of features of each individual over consecutive time segments in the media timeline. This enabled comparisons between subject groups to be made on the basis of SCR score (higher score with more engaging event) and SCR amplitude (higher amplitude with a deep level of engagement).
The same level of flexibility in the selection of sensing equipment was not present in Study III as there were many artefacts due to cockpit motion and bodily movements while driving (discussed under Challenge 3 of Chapter 3). Although the researcher selected EEG to overcome most of the drawbacks, there were several other challenges such as the uncertainty of reliability of EEG equipment (see Appendix B.2) and the difficulty of analysing EEG (most researchers including Nählinder, 2009 have avoided EEG labelling it as difficult to analyse), however, this research proposed an improved computational technique for analysing EEG data by combining EEG-based vigilance estimators, performance measures, and subjective ratings (see Appendix B.1). The proposed method helped to generalise the results over different groups of individuals and experimental conditions in simulator validation.

Apart from the psychophysiological measures discussed above, there were several other measures/features (e.g. heart rate variability and EEG-based event related potentials) that did not succeed as generalised results for various reasons (e.g. unreliable equipment and higher signal-to-noise ratio). Those measures were excluded from further analysis.

RQ3:

How do different measures of behaviour reflect the body’s role in cognition/behaviour?

According to the assumption of psychophysiology (see Section 1.1), the capacity to reflect the processes of the human mind in psychophysiological measures confirms the embodied nature of human cognition. In this research, Study I identified certain patterns of psychophysiological measures depending on the type/level of engagement of participants, however, embodied cognition is more radical than such simple relationships between psychological processes and physiological events (see Chapter 2). This research thus attempted to uncover more advanced relationships so as to understand the body’s role in cognition/behaviour using the three types of measures incorporated, as discussed below.

The general implication of Paper IV (see Study I) is that the depth of engagement is higher in scenarios based on real actors than in those based on animated characters, however, the researcher identified a deficiency in this conclusion as there is uncertainty over the influence of the sound track. Participants who watched the animated character film may have experienced cognitive disequilibria due to certain mismatches between auditory and visual content, as well as low quality animations. In other words, the absence of a deep engagement in the animated film may have resulted from the lack of skill perceived by participants (see Noë, 2004 for his conception of perception as a kind of skilful bodily activity).
The methods proposed in Study II are based on the assumption that the psychological processes of a trainee determine their external behaviours, which can be recognised using performance measures. For instance, the method proposed in Paper III is based on the assumption that a participant is motivated to behave in a certain way as a result of the type of goal adopted. If this is the case, all three types of measures should converge to give a unified understanding of the behaviour of a subject. For instance, in a subject who behaves in a certain way on the bends of a road to make turns, there should be a relationship between the measures such that while performance measures reflect the preparation or the performance during turns, the psychophysiological measures should reflect the internal processes which support those external behaviours. Although this research has analysed data using a number of techniques (e.g. event-related potentials and time-frequency analysis of EEG; see Challenge 5 in Chapter 3), no such relationship has been identified between those measures (however, several controlled laboratory experiments have reported certain relationships in driving situations, e.g. Haufe et al., 2011). The absence of such simple relationships in unconstrained driving tasks relates to the embodied nature of an interaction. In other words, the mind is not the sole influencer of cognition or behaviour, rather it is determined by the interaction between mind, body, and environment (according to Gibson (1958), an individual’s perception-action patterns are constantly affected by the invariant information of the situated environment).

In Study III, there were several results relating to the body as a constraint on the cognition thesis of embodied cognition (see Chapter 2). For instance, the researcher expected a similar (or higher) level of performance as in real world driving from expert drivers when they drove in the driving simulator, however, both performance and subjective measures demonstrated that the experts experienced a certain difficulty in translating their skills in the simulator. Apart from the above, the results of both simulator and real world driving showed that driving behaviours are highly task dependent and depend on the characteristics of drivers, such as their age and past driving experience (similar results have been reported in the literature, see Mullen et al., 2011; Ouimet et al., 2011).

RQ4:

*How can an indirect approach, using performance, psychophysiological, and subjective measures, be utilised to validate the equivalence between simulator and real world training?*

This research answered the above research question in several stages. First, a comparison was performed in Study I using psychophysiological measures to evaluate whether there was a degradation in engagement to see-
narios based on animated characters compared to those with real actors (here it was assumed that there is a correspondence between scenarios based on animated characters and simulated scenarios, as well as between real actor based scenarios and real world scenarios). Although the results of Study I was in favour of a weaker degree of engagement with animated character scenarios with compared to real actor scenarios, the researcher raised a concern about this implication (see Study I; also elaborated in Section 5.3). Training scenarios are usually different from the scenarios used in Study I as, they require a participant to respond. The scope of Study I was thus limited by such elements.

Study II, by fixing certain drawbacks of Study I (e.g. capturing the response of participants), proposed two methods to distinguish the behaviours of trainees with respect to their training domain based on both subjective psychological elements and performance measures. Although the proposed methods have advantages in the assessment of training, this research did not use those methods for the validation of training simulations. The reasons for this decision included the researcher’s view that neither subjective and performance measures alone can give a reasonable understanding of human behaviour and the perceived complexity of an experiment in real world.

Finally, Study III eliminated most of the limitations that were present in the other two studies and focused specifically on the validation of a driving simulator using the three types of measures (see Creswell (2013), for embedded mixed methods design). It used an experiment design inspired by both backward transfer and quasi-experimental studies as well as naturalistic experiments (see Section 1.3 on Research Approach and Challenge 7 of Chapter 3). The design of the experiment also used existing infrastructure (e.g. driving simulator and sensing equipment) and participants (e.g. driving instructors to perform both on-road and in simulator driving tasks). The comparisons between different conditions of the experiment (e.g. contexts, subject groups, and driving tasks) were made by examining the measures of relative validity (see Caird & Horrey, 2011), mental states (e.g. vigilance), and participant opinions about different elements.

5.3 Conclusions about the research problem

The previous section discussed the conclusions about the methodological elements of the research problem which were distributed across the four research questions. This section focuses on the overall theoretical conclusions about the research problem.

This research addressed an enduring research problem in simulation research: the validity of training simulations. To address this problem, this study adopted an indirect approach using three types of behavioural measures. This also included identifying the impact of virtual reality training
environments on the nature of engagement of participants, and the implications of the training effectiveness of simulations, however, as this research has infused insights of embodied cognition in its conceptual framework, interpretations of its findings are somewhat different from the results of most existing studies.

One of the major implications of this research is the existence of different types of participant engagement depending on the uniqueness of situations. The factors or variables that influenced the uniqueness of a situation have been identified as the type of context, the nature of task, the level of task-related proficiency, and the age of an individual (Mullen et al., 2011 discuss a similar list). For instance, the research revealed that the engagement with a scenario based on animated characters was different from engagement with a real actor scenario. This research identified certain other results for differentiating real world driving from a driving in a simulator. For instance, expert drivers experienced certain difficulty in adapting to the simulator or translating their driving skills during their preliminary exposures in the simulator, however, novice drivers were able to adapt in the simulator much faster than experts. Although some of these results agree with the perception that a carefully designed training simulation (e.g. having an appropriate psychological fidelity) has the potential to offer a real world-like training experience, a further inference from the results raises concerns about this. The source of this disparity is the higher level of vigilance of participants during simulator driving. This result can be interpreted considering several other elements as discussed below.

A number of authors have reported the importance of appropriate mental conditions or states for learning, such as a moderate level of arousal, positive emotions, and higher degree of attention (see Biggs & Tang, 2007, p. 31; Carini et al., 2006; Galotti, 2004; Gibbs & Habeshaw, 1992; Picard et al., 2004). A higher level of vigilance in the participants is thus in favour of the conclusion that the driving simulator is an effective training tool, however, this should be questioned from two other perspectives. Is the increased vigilance a symptom of an entertainment oriented engagement of participants, and to what extent are the skills acquired in a simulator transferable to the real world if psychophysiological signatures are different between the two contexts? There are several results to support the former, including the increased performance of novice drivers and the adaptive behaviour (including the flow experience) of expert drivers. Another facet of this behaviour can be understood as follows. In the real world, most drivers use their own vehicles and they have a purpose for driving, e.g. to go to an office, however, it is not the case in a simulator and, as a result, the participants are either free to adopt a trip purpose themselves, or they have to do so under an external influence (e.g. as specified in driving lessons). This can ultimately lead to the entertainment-oriented engagement of participants. Although some advocates promote entertainment in CAI (see discussion about serious games in
Section 2.3.3), it can have negative effects, especially on the perception of risk (Ranney, 2011).

The latter problem is more serious than the former as it determines the depth or complexity of the driving skills acquired from simulator-based training. This can be explained using an example. Consider a real world crash avoidance situation. In such a situation, depending on the past driving experience and abilities of the driver, they detect the hazard a few seconds (or milliseconds) before the crash, which is also associated with an action pattern to cope with the situation. According to theories in contemporary cognitive science (e.g. fight-or-flight behaviour and enactive cognition), the behaviour of the driver during such situations cannot be explained based on the cognitive elements of driving skills alone, rather it depends heavily on unconscious and body-based skills in which psychophysiological signatures play a major role. The observed difference of vigilance between simulator and real world driving thus has negative consequences on the validity of the simulator.

After considering the results showing both positive and negative support for the validity, this research arrives at the conclusion that training simulators have a limited validity in the training of skills. While some of these limitations can be overcome through a proper combination of physical and psychological fidelity, there is another category of limitations which cannot be fulfilled unless tested in an operational context (real world), however, in the middle, there are certain limitations that can be overcome by focusing on psychophysiological signatures and improving them so as to be more realistic in the real world.

5.4 Contributions

This research makes both epistemological and methodological contributions in simulation research which can be discussed as attributed to the three themes of studies presented in Chapter 4. Study I demonstrated how psychophysiological measures can be used to determine the nature and depth of the engagement of participants, and to use the developed methodological framework to compare scenarios using animated characters and those using real actors. Apart from the findings of Study I (already discussed in previous sections), its contributions included mapping different classes of engagement and associated patterns in the feature of GSR signals and an identification of two types of GSR features (i.e. mean SCR score and mean SCR amplitude) for comparisons of user engagement in different types of scenarios. The proposed methods are more reliable and advanced than most existing methods. For instance, although most studies have used subjective measures such as ratings from questionnaires and facial expressions, they have disadvantages, including subjective biases, and they do not provide a continuous measure of
engagement (Fairclough, 2007; Picard et al., 2004). Most previous studies based on GSR (e.g. Wang & César, 2015) have relied on basic tonic level changes rather than advanced phasic level changes of GSR (see Cacioppo et al., 2007, for details).

The contributions of Study II included proposing two methods to quantify and differentiate the behaviours of trainees with respect to their training domain, which is a prerequisite in simulation validation. The proposed methods also provide an indirect way of assessing the skills of trainees, especially when they engage in experience-based learning (EBL) activities such as driving. The most frequently used assessment methods in EBL, such as observation by instructors, after-action reviews, and experience-points, have disadvantages, such as being labour and time intensive, subjective biases, and evaluating the outcome of training as a product rather than a process (Bayliss & Schwartz, 2009; Drews & Bakdash, 2013; Liu et al., 2008; Picard et al., 2004; Wiese et al., 2008), however, the proposed methods allow the training to be monitored as a continuous variable in an objective way, which is more beneficial in systems which can enable an adaptive task challenge (see the discussion about flow theory in Chapter 2).

Finally, Study III resulted in several major contributions to simulation research, including more advanced insights into both psychophysiological measures and embodied cognition for simulation validation. As discussed in this thesis (e.g. Section 1.3), most previous studies which have used behavioural measures have analysed each data type in isolation from the others (e.g. Backlund, Engstrom, Johannesson & Lebram, 2008; Nählinder, 2009), have not considered the embodied nature of human cognition/behaviour, and have avoided naturalistic experiments. Previous studies have avoided EEG labelling it as difficult to analyse/interpret (Nählinder, 2009), however, this research has overcome the drawbacks of previous studies by incorporating EEG data and successfully combining it with performance and subjective measures (see the answers to RQ1 & RQ2 in Section 5.2). Apart from the methodological contributions, Study III makes major theoretical contributions about the validity of training simulations, as discussed in the previous section.

5.5 Limitations

This section presents certain important limitations of this research that became apparent during its progress, however, it should be noted that this research has been conducted in an incremental manner while going through several cycles in the research ‘wheel’ (Figure 7), and by addressing most obvious limitations/drawbacks of previous stages as explained in the research process (Section 3.3) and under each study (Chapter 4).
From the beginning, this research suffered from a major limitation as to supplying relevant subjects in satisfactory numbers for its empirical work. As a result, some parts of this research were not thoroughly validated (especially Study II) and others were conducted under certain assumptions, however, justifications were provided in such situations to mitigate those limitations (e.g. driving expertise based on the number of years of active driving).

This research also had a limitation in its equipment, especially the psychophysiological sensing equipment (e.g. higher noise ratio and limited number of electrodes). If more effective equipment had been used, there may have been opportunities to obtain more advanced results, such as about the roles of the motor and somatosensory cortex during training and ERP based results.

Another limitation of this research is the absence of certain performance measures and judgments by domain experts. For instance, in Paper III, the qualitative assessment was performed by the researcher of this thesis rather than involving a driving instructor.

Several other limitations were associated with the research design and data analysis. For instance, in Study III, ratings of the simulator experience were recorded at the end of each session rather than at the end of each task, which ultimately limited the scope of data analysis. This research has omitted the analysis of certain data, such as the responses to open-ended questions and certain sophisticated data analysis techniques, such as the Granger causality test for practical reasons.

Finally, some of the delimitations defined at the beginning of this research (Section 1.5) have also created circumstances for certain limitations. For instance, although this research decided to drop the entertainment element from its main considerations (as it did not focus on serious games), the results and observations prompted the importance of considering that aspect as an inherent attribute of users of simulations.

5.6 Implications for further research

This research has focused specifically on one of the training domains (i.e. driving) and a category of skills. Further research could be initiated considering other types of skills and domains of training. Another important direction for further research is by removing some of the limitations mentioned in Section 5.5. For instance, if more effective sensing equipment and analytical tools are used, there is an opportunity to capture ERPs for different types of events (e.g. incidents) during the interactions of trainees in simulators, so as to study brain activity patterns (a limited study in this direction was conducted, as discussed, in Appendix B.2). This research needs further validation/enhancement/refinement for its proposed methods/techniques (e.g. the formula-based approach for assessing performance competence) both from
theoretical and empirical standpoints. Finally, there are many other application domains for which the methodological framework proposed in this research can be adopted, such as adaptive games, e-learning, and mind machines.
References


http://doi.org/10.1162/dmal.9780262693646.021


Wise, & M. Mouloua (Eds.), *Human Factors in Simulation and Training* (pp. 49–60). London: CRC Press.


Appendix A: Publications

*Paper I: Determining the Psychological Involvement in Multimedia Interactions*

*Paper II: Game Interaction State Graphs for Evaluation of User Engagement in Explorative and Experience-based Training Games*

*Paper III: Assessing Performance Competence in Training Games*

*Paper IV: Affective Realism of Animated Films in the Development of Simulation-Based Tutoring Systems*

*Paper V: Comparing Expert Driving Behaviour in Real World and Simulator Contexts*

*Paper VI: Comparing Expert and Novice Driving Behaviour in a Driving Simulator*