Interaction Design and Architecture(s) Journal - IxD&A, N.19, 2013, pp. 115-131

Comparing Expert and Novice Driving Behavior in a Driving Simulator

Hiran B. Ekanayake[1,3], Per Backlund[2], Tom Ziemke[2], Robert Ramberg[1], Kamalanath P. Hewagamage[3], and Mikael Lebram[2]

[1]Stockholm University, Forum 100, SE-164 40 Kista, Sweden
[2]University of Skövde, P.O. Box 408, SE-541 28 Skövde, Sweden
[3]University of Colombo School of Computing, 35, Reid Avenue, Colombo 7, Sri Lanka
hbe@ucsc.lk, {per.backlund, tom.ziemke}@his.se, robban@dsv.su.se, kph@ucsc.lk, mikael.lebram@his.se

Abstract. This paper presents a study focused on comparing driving behavior of expert and novice drivers in a mid-range driving simulator with the intention of evaluating the validity of driving simulators for driver training. For the investigation, measurements of performance, psychophysiological measurements, and self-reported user experience under different conditions of driving tracks and driving sessions were analyzed. We calculated correlations between quantitative and qualitative measures to enhance the reliability of the findings. The experiment was conducted involving 14 experienced drivers and 17 novice drivers. The results indicate that driving behaviors of expert and novice drivers differ from each other in several ways but it heavily depends on the characteristics of the task. Moreover, our belief is that the analytical framework proposed in this paper can be used as a tool for selecting appropriate driving tasks as well as for evaluating driving performance in driving simulators.

Keywords: Educational game design, simulation-based training, playful learning, emotion in games, driving simulator, simulator validity, evaluation, driving performance, psychophysiology, EEG, Emotiv EPOC, player experience.

1 Introduction

Motor vehicle crashes and fatalities are among the highest ranked cause of deaths worldwide. For instance, according to U.S. department of transportation, 33,561 people have lost their lives on roadways because of crashes in 2012 [1]. According to literature, about 20-30% of crashes occur as the lack of capability of drivers to orient their attention on such situations because of bad vigilance such as boredom [2], [3]. This prompts the need for an additional learning experience other than the driving skills required to maneuver a vehicle in different road traffic and environmental conditions for which driving simulators are the ideal experiential learning environments [4], [5].

Much research reports on the successful use of driving simulators to enhance learning. For instance, a game-based driving simulator has been successfully used to improve traffic safety variables such as speed, use of turn signals and rear-view mirrors, headway distance, and lane change behavior [6]. Some other advantages of using driving simulators are faster exposition to a wide variety of traffic situations, feedback from different perspectives, unlimited repetition of educational situations, automated and objectivistic assessment, demonstration of maneuvers, analyze risky scenarios without endangering participants, and controllable and adaptive learning tasks [7], [8], [9]. Although simulators are increasingly used for training purposes, there is a question of training effectiveness of simulations, that is, the knowledge about the transferability of performance and competence achieved in a simulator when it is applied in real world. According to literature, numerous studies have addressed this problem, for example [6], [9], [10], and [11]. However, still simulation research lacks good theory-based approaches or the use of academically acceptable methodologies for evaluating the training effectiveness of simulations [4], [12].

According to Feinstein and Cannon [12], this problem stems from there still being a disagreement among researchers about measures that can effectively validate a simulation or the required level of fidelity in a simulation-based training environment. The validation should necessarily evaluate the accurate (algorithmic) representation of a desired phenomena, participants' perception about the phenomena being modeled, and the phenomena's relation to the real world situation being modeled. However, a validation focusing only on the above mentioned aspects would become meaningless unless the desired output of an educational simulation is not taken into consideration. For instance, although high fidelity simulators are capable of rendering a higher degree of realism, it can actually interrupt the learning experience by over stimulating novice trainees. The same issue has been raised by Bell and colleagues [4] from a different perspective and their suggestion is that designers should focus on the instructional features first and technology choices as the end point of the design process. These problems are also faced by designers of so called serious games in their effort to balance three aspects in a serious game, which are, intrinsically motivating game play, immediate feedback, and the learning impact [13, p.5].

Considering the importance of the problem, we previously conducted a study in which we compared driving in the real world with in a driving simulator using both performance and psychophysiological measures [14]. However, in that study we did not compare the difference between novice and expert driving behaviors in order to derive implications of learning effects of novice drivers in the simulator context. To this end, our present study has compared the driving behavior of expert and novice drivers in a mid-range driving simulator by analyzing and triangulating various aspects of driving behavior. Although our approach is mainly driven by data and inductive reasoning, we deem that our approach is methodologically sound and valid as most scholars have justified the use of triangulation of data to strengthen the evidence as well as raised the importance of inductive reasoning for extending current knowledge boundaries [15], [16].

The paper is organized as follows. Section 2 discusses three commonly used measures in for evaluating simulator effectiveness. Section 3 presents the method which includes the justification of our approach, the experimental setup, and the procedures we used for analyzing different types of measures. The results are presented in Section 4 and the findings are discussed in greater detail in Section 5. Finally, Section 6 includes the conclusions of our study.

2 Validating Simulation-Based Training

Although there are numerous theories about experiential learning such as Kolb's experiential learning cycle [17], there are still difficulties formulating an effective relationship between action (experience) and knowledge (conceptualization). This ultimately makes it hard to determine adequate assessment strategies as well as to 'prove' a simulator's effectiveness for training [18], [19]. Therefore, taking in-situ action as the expected outcome of an effective training situation, many researchers report on approaches that are mainly driven by data to validate training effectiveness of simulations.

According to literature, there are three different measures for evaluating the equivalence of driving between different groups and contexts, that is, performance, psychophysiological, and subjective measures [10], [11], [20], [21], [22], [23], [24]. Performance measures monitor physical and behavioral changes (e.g., lane changing behavior, vehicle speed, head movements, and steering wheel variance) and capture how well the user is performing a given task. Psychophysiological measures are the resulting physiological changes such as heart-rate variability (HRV), galvanic skin response (GSR), and electroencephalographic (EEG) signals for psychological manipulations, and it covers a broad range of aspects of perception and attention and related processes. However, it is very difficult to interpret psychophysiological indices as it often contains many-to-many relationships between psychological elements and physiological elements [25], [26]. Finally, subjective measures are based on self-reported measures of user experience which is usually captured using questionnaires and interviews. However, subjective measures are considered problematic because of the unreliability of self-reported emotional information and the requirement to interrupt user experience [27].

Numerous studies (e.g., [6], [9], [10], and [11]) have evaluated training effectiveness of simulations using one or a combination of the above measures. However, those studies seem to have omitted either direct comparisons between driving in the real world and driving in a simulator or triangulation of different measures to enhance the validity of results.

3 Method

3.1 Our Approach

As described in the introduction section, we are primarily using a data driven approach with triangulation of different measures to evaluate the training effectiveness of simulations. For this validation, we identified two options, that is, either to test novice drivers in the real world after having trained in a simulator or to compare novice drivers against expert drivers in the same learning context (the simulator). However, after considering the ethical side of exposing novice drivers to drive in real world (see [24]), we decided to go with the latter. Therefore, our approach is based on comparing novice drivers' driving behavior with that of expert drivers in a driving simulator using various measures. This approach is basically driven by the assumption that expert drivers behave similarly (i.e., their performance including decision making) across the two contexts (see [28] for a discussion of expert level of skills). For selecting expert drivers, we relied on their driving experience, that is, the number of years of driving after obtaining the driving license as well as the ability to instruct others about driving.

For the experiment, we employed a mid-range driving simulator. Although it is not a high fidelity simulator, we deem that the technological quality of the simulator was sufficient as it can imitate both physical behavior and scenarios to a satisfying degree for training tasks (see [12] for a discussion about the need for a balance between realism and instructional capabilities of a simulation). As the measures of our experiment, we collected data about several performance, psychophysiological, and subjective variables. However, rather than relying on the results of such measures separately, we tried to combine different measures such as by using correlation analysis to enhance the reliability of our findings.

Apart from other types of measures, we considered EEG based vigilance estimators as a way to look into how higher cognitive processes are involved with driving. Vigilance refers to the ability of organisms to maintain their focus of attention and to remain alert to stimuli over prolonged periods of time [2], and in the introduction section we discussed the importance of driver vigilance and its relation to road fatalities. The complex relationships between driver vigilance, road scene, vehicle speed, peripheral vision, and mental workload are discussed in [20] and [23].

3.2 Equipments and Tools

The main equipment of the experiment, the driving simulator, comprised of a real car surrounded by seven screens covering the whole field-of-view for the driver, including the parts covered by the rear-view mirrors (220x30 degrees forward and 60x30 degrees rear). Both sound vibrations and the car's fan helped to create an illusion of movement (see [6] and [8] for more details about the driving simulator). The physical performance data were gathered from the game engine itself (e.g., speed of the vehicle) as well as by attaching relevant sensors (e.g., linear potentiometers) to brake and gas pedals and steering wheel and by sampling the readings at a rate of about 100 S/s using a microcontroller module and feeding the data to the PC. Moreover, two cameras provided the frontal field-of-view and view of the driver. For generating sceneries and relevant physical behavior we employed two different game engines – VDrift (http://vdrift.net/) and OGRE (http://www.ogre3d.org/) – which come with realistic physics.

For capturing psychophysiological measures, we employed the Emotiv EPOC neurofeedback headset (http://www.emotiv.com/). The headset uses 14 sensors and two references to capture EEG potentials from several important scalp locations as well as providing two-axis gyro data for detecting head movements. There are several benefits of using the Emotiv headset such as low cost, wireless communication, and it is easy to setup. However, it has certain limitations including lower sampling rate, high signal-to-noise ratio, and it does not cover some important scalp positions. Nevertheless, the headset was ideal for our experiment as it is relatively unobtrusive (see [29] for a discussion of sensors for highly dynamic environments) and we were able to eliminate its limitations to a greater degree by accessing raw EEG data and focusing on slowly varying emotional information rather than time-locked EEG activities (see Section 3.5). The headset has been used extensively for research purposes (see http://www.emotiv.com/ for a list of published papers which have used the EPOC headset).

We primarily used Matlab [30] and EEGLAB [31] for data analysis including segmenting, re-sampling, interpolating and smoothing of data, filtering EEG data into different bands, and obtaining certain graphs. For comparing means of different groups balanced one-way ANOVA (Analysis of Variance) was used. ANOVA offers a greater flexibility for comparing means of more than two groups which is not possible with Student's t-test [40, p.115].

3.3 Participants

A total of 31 healthy participants (mean age = 26.7 years and SD = 12.9 years; 24 male and seven female) took part in the experiment after providing a written statement of informed consent. The participants were recruited within three driver categories: driving instructors from a well known driving school (N = 8; 27-56 years; mean age = 40.9 years and SD = 11.5 years; five male and three female), regular drivers within the university staff (N = 6; 26-51 years; mean age = 36.7 years and SD = 8.4 years; three male and three female), and novice drivers from another driving school (N = 17; 16-18 years; mean age = 17.0 years and SD = 0.4 years; 16 male and one female). Both driving instructors and regular drivers had extensive driving experiences (mean driving years of 23.6 and 18.2, respectively). After each experiment, each participant received a free lunch and refreshments as compensation for their involvement in the experiment.

3.4 Driving Tasks

In the driving simulator, expert drivers participated in two driving sessions, whereas novice drivers participated in up to three sessions. As our previous study [14] revealed that there is no substantial difference between the driving behaviors of driving instructors and regular drivers, we grouped both of them as expert drivers. In each driving session, they drove in the OGRE based highway traffic track having levels of increasing difficulties for about 10 minutes; in VDrift Monaco track (city area like track, but no traffic) for about 5 minutes; and finally in VDrift LeMans track (landscape like track, but no traffic) for about 5 minutes. Fig. 1 shows screenshots of the tracks of the simulator driving session.

(a) (b) (c)

Fig. 1. Screenshots of (a), OGRE based highway traffic track; (b), VDrift Monaco track; and (c), VDrift LeMans track.

The tasks were labeled using a naming convention of the form <driver category: E-expert and N-novice>-<driving session: 1, 2 or 3>-<driving track: T30-Highway, T41- Monaco, T51- LeMans>. For example, 'E-S1-T30' refers to 'Expert drivers - Session 1 - Highway traffic track'. Each subject completed a questionnaire in a quiet office soon after each session of driving. In the questionnaire each subject had to answer questions about their driving experience, disturbances, and several other aspects.

3.5 Data Analysis

For analyzing the data, we employed the same analytical framework we proposed in our previous study [14] but we extended it further to incorporate qualitative measures. In the analysis, we identified four types of performance measures, that is, speed of the car, steering wheel angle, gas pressure, and brake pressure. After preconditioning the data (e.g., removing noise and fixing discontinuities), we derived eight variables from those measures, that is, means of speed, means of steer, means of gas, means of brake, SDs of speed, SDs of steer, SDs of gas, and SDs of brake. Values for the abovementioned variables were calculated in the following manner, for instance, means of speed of a driver group was calculated by averaging each member driver's mean speed values of a given driving track whereas SDs of speed of a driver group was calculated by averaging each member driver's standard deviation of speed values of a given driving track.

EEG data were analyzed in the following way. First, for each channel data of each EEG recording we obtained the band powers for consecutive one-second durations of seven frequency bands, that is, delta (1-4Hz), theta (4-7Hz), alpha1 (7-10Hz), alpha2 (10-13Hz), beta1 (13-22Hz), beta2 (22-30Hz) and gamma (30-45Hz). The decision to categorize the EEG spectrum to different frequency bands was motivated by literature [26], [32] which suggests that different frequency bands are associated with different states of the mind, that is, delta with deep dreamless sleep, theta with light sleep or impaired information processing, alpha with awake but relaxed state without processing much information, beta with the normal state of wakefulness with increased cognitive activity and attention, and gamma with various types of learning and higher cognitive processing. After the categorization, to eliminate noise from each component, we used the gyro x and gyro y data whose signal magnitudes exceed a threshold (i.e., 3*standard deviation) and linear interpolation to fill up the gaps. Next, each band power component was smoothed using the loess algorithm (local regression using weighted linear least squares and a second degree polynomial model) with a 30 second time span. According to literature (e.g., [21], [33], and [34]), EEG features with minute-scale smoothing can be used as vigilance estimators. After this step, each vigilance component was processed to find up to six peaks, which were highest, and up to six valleys, which were lowest, within the waveform. At the same time corresponding (within the same timeline) values of each driving variable (i.e., speed, steer, gas, and brake) were recorded. Next, features were filtered by performing ANOVA F-tests to identify whose means were significantly different between a particular driving variable's values at peaks and valleys ($p < 0.05$). Finally, we averaged the values over different frequency bands and used as measures to compare between the driving behaviors of different groups and tasks. For example, Figure 2 contains a graphical representation of the values obtained for peaks and valleys as well as mean values and standard deviations of speed. To quantitatively determine the features of the graphs, we obtained two types of measures called VG and TD, which were calculated as follows:

$$\text{VG} = \text{means of X} - \text{mean X at valleys} . \tag{1}$$

$$\text{TD} = \mid 2^*\text{standard deviation of X} / (\text{mean X at peaks} - \text{mean X at valleys}) \mid . \tag{2}$$

Where X is a performance variable.

The measure VG (1) determines the relationship between a performance variable and driver vigilance. For instance, if the VG measure is negative between speed and vigilance, then it implies that driver vigilance gets increased when the driver drives in low speeds.The measure TD (2) is associated with the task demand of a variable. That is, if TD is less than one it implies that the task demand is lower than the comfort range; otherwise the task demand is higher (we refer to comfort range as the zone between peaks and valleys of vigilance associated with a performance variable).

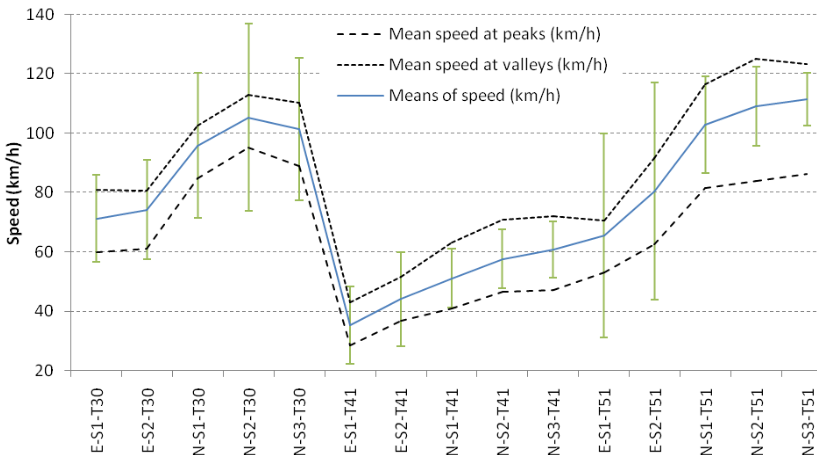

Fig. 2. Graphical representations of the values at peaks and valleys of EEG vigilance estimators and means and standard deviations (as error bars) of speed.

The user experience questionnaire contained the seven dimensions of player experience questionnaire, commonly known as the In-Game Experience Questionnaire (iGEQ), and several other questions to rank the disturbances from the experimental setup and certain other aspects. The iGEQ measures a user's game experience within seven dimensions, that is, Immersion, Tension, Competence, Flow, Negative Affect, Positive Affect, and Challenge [35], [36]. Each dimension in the iGEQ is formed by two questions, e.g., tension using "I felt frustrated" and "I felt skillful," and uses a five point Likert-type scale for ranking. Three additional questions were included to test their satisfaction of driving, self-evaluation of their driving skills improvement, and their opinions on using the simulator for driver training. Table 1 presents the questions comprised of iGEQ dimensions and the three additional questions.

Table 1. Some questions in the user experience questionnaire.

Question	iGEQ dimension
I was interested in the game's story	Immersion
I felt successful	Competence
I felt bored	Negative Affect
I found it impressive	Immersion
I forgot everything around me	Flow
I felt frustrated	Tension
I found it tiresome	Negative affect
I felt irritable	Tension
I felt skillful	Competence
I felt completely absorbed	Flow
I felt content	Positive affect
I felt challenged	Challenge
I felt stimulated	Challenge
I felt good	Positive Affect
How much are you satisfied with your driving skills, during your last driving session?	NA
Have you improved your driving skills during your last driving session?	NA
Do you recommend the driving simulator for training drivers?	NA

To check whether self-reported measures of player experience correlates with psychophysiological measures we have followed the following procedure. After the step that obtained up to six peaks and six valleys for each vigilance component described above, the features were filtered based on means that were not significantly different from each other between peaks and valleys (P < 0.05). Next, values of peaks and valleys were averaged over each frequency band as well as differences between the averages were also recorded. Finally, Pearson's correlation coefficients were calculated between iGEQ dimensions and each set of EEG based vigilance estimators.

4 Results

As the first step, it was required to check whether different conditions of the experiment, that is, driver type, driving session, and driving track, have affected the readings differently or to what degree. For testing this aspect a multiway ANOVA F-test was performed (Table 2) involving the eight performance variables and considering only those tasks of which the other conditions were stable among both groups.

Table 2. Multiway ANOVA *F*-test values for testing the effects of multiple factors.

Variables	Means of speed	Means of steer	Means of gas	Means of brake	SDs of speed	SDs of steer	SDs of gas	SDs of brake
Driver type F(1,160)	53.9***	45.6***	52.4***	29.8***	40.3***	48.1***	27.0***	30.5***
Driving session F(1,160)	5.66*	4.57*	4.55*	2.43	5.77*	1.39	4.67*	2.90
Driving track F(2,160)	63.2***	143***	201***	15.1***	15.3***	87.3***	19.0***	6.92**

*, **, *** significant differences at P < 0.05, P < 0.01, and P < 0.001, respectively.

According to Table 2, there are significant main effects on driver type and driving track (P < 0.001). However, the mean effect on driving session is less significant (P < 0.05) and limited to a few variables. To make our analysis consistent with our previous study [14], we have considered only three variables for further analysis, that is, means of speed, SDs of steer, and means of gas. However, before proceeding further, we have compared several other conditions such as disturbances from the mere presence of researchers and different types of equipment that may have influenced the subjects differently during different experimental conditions. Yet, we found that none of the predicted sources have affected the experimental conditions to a substantial degree as both means and standard deviations were very low in each case.

4.1 Comparing Driving Behavior of the Two Driver Types Based on Performance Measures

The graphs were obtained for the three variables considered over different driver types, driving sessions, and driving tracks (Figure 3).

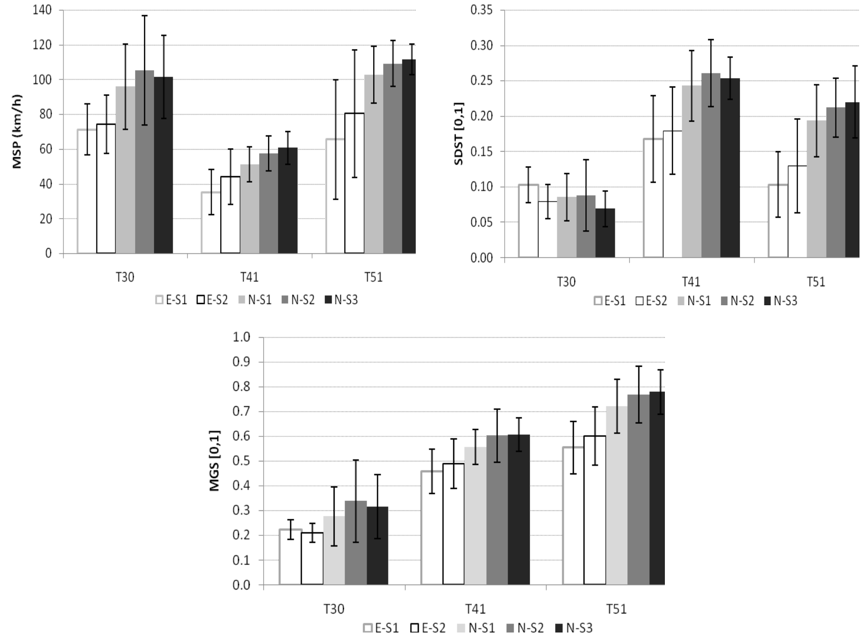

Fig. 3. Values of means of speed (MSP), SDs of steer (SDST), and means of gas (MGS) of expert (E) and novice (N) drivers in different sessions and tracks of driving in the simulator and standard deviations based error bars.

According to Figure 3, mean speed values have increased during successive sessions (except N-S3-T30) and the values are higher in expert drivers than in novice drivers. Moreover, in T30, standard deviation values of speed of expert drivers are lower than novice drivers, whereas, in T41/T51, the order is reversed. A similar pattern can be observed with respect to MGS which some exceptions (i.e., E-S2-T30 and N-S3-T30). The SDST values also show somewhat similar pattern as of the other two variables, again with some exceptions (i.e., E-S2-T30, N-S3-T30, and N-S3-T41).

4.2 Comparing Driving Behavior of the Two Driver Types Based on Both Performance and Psychophysiological Measures

The results of the combined analysis of performance and psychophysiological measures described in Section 3.5 are graphically represented in Figure 4.

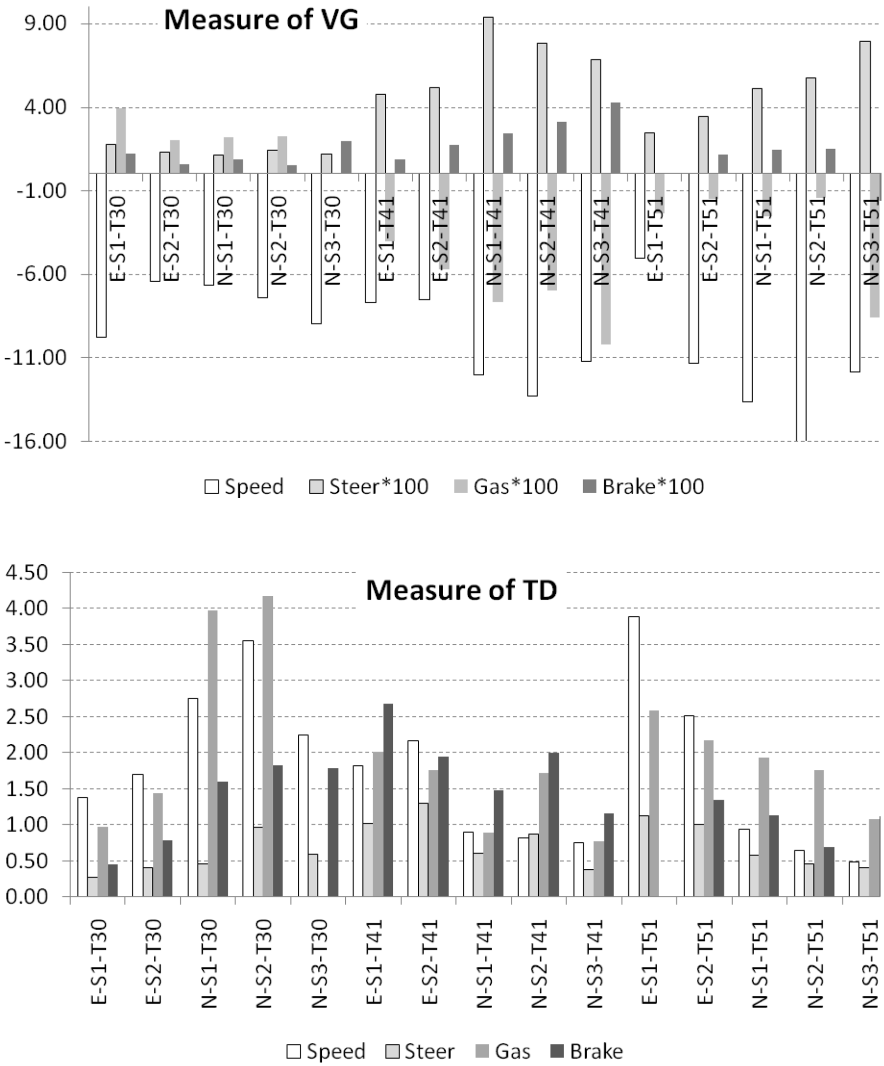

Fig. 4. Graphical representations of the measure of VG (top) and the measure of TD (bottom) over different driver types, sessions, and driving tracks

The measure of VG represented in Figure 4 can be summarized as follows. Driver vigilance gets increased under low speed driving, increased steering, or increased braking. However, gas causes to increase driver vigilance in T30 but not in T41/T51. In T41/T51, driver vigilance caused by steering has affected novice drivers more than expert drivers. However, in T41/T51, driver vigilance caused by speed has affected expert drivers more than novice drivers.

Apart from the above, according to the graph of the measure of TD in Figure 4, T30 is more task demanding for novice drivers as compared to expert drivers.

However, in T41/T51, the effect is almost the opposite as the tracks are more task demanding for expert drivers as compared to novice drivers. Moreover, the tracks T41/T51 seem to be more comfortable for novice drivers as TD values are less than one of both speed and steer variables.

4.3 Comparing Driving Behavior of the Two Driver Types Based on Self-Reported Measures of User Experience

The results of the In-Game Experience Questionnaire (iGEQ) as well as three additional questions of the self-reported questionnaire are given in Table 3.

Table 3. Means and standard deviations (within parentheses) of iGEQ dimensions and other questions of different groups and driving tasks.

Variable	E-S1 n=14	E-S2 n=13	N-S1 n=18	N-S2 n=15	N-S3 n=6
[iGEQ] Immersion	2.3 (0.8)	2.3 (0.9)	2.6 (1.0)	2.1 (0.9)	2.3 (1.1)
[iGEQ] Competence	1.6 (0.8)	2.2 (0.8)	2.1 (0.5)	2.3 (0.6)	2.7 (0.7)
[iGEQ] Flow	2.4 (0.9)	2.1 (0.5)	1.7 (1.0)	1.5 (0.9)	1.7 (1.3)
[iGEQ] Tension	1.8 (1.1)	1.1 (1.0)	0.6 (0.6)	0.6 (0.8)	0.1 (0.2)
[iGEQ] Negative affect	1.1 (1.0)	1.1 (1.2)	0.6 (0.6)	0.6 (0.6)	na
[iGEQ] Positive affect	1.9 (1.0)	2.6 (1.0)	3.0 (0.6)	3.0 (0.5)	3.2 (0.9)
[iGEQ] Challenge	2.3 (0.8)	2.5 (0.9)	2.8 (0.8)	2.2 (0.8)	2.4 (0.7)
Satisfaction of driving	1.5 (0.9)	2.7 (0.6)	2.2 (0.8)	2.7 (0.9)	2.7 (1.0)
Whether skills improved?	0.8 (1.3)	2.4 (1.5)	1.3 (1)	1.7 (0.9)	2 (1.3)
Whether recommending the simulator for driver training?	1.7 (1.1)	2.6 (0.9)	1.9 (1.3)	1.9 (1.1)	2 (1.3)

The pair-wise comparison of means of the seven iGEQ dimensions between the pairs (E-S1, E-S2), (E-S1, N-S1), (N-S1, N-S2), (E-S2, N-S2), and (N-S2, N-S3) reported that statistically significant differences of means exist in tension, negative affect, and positive affect dimensions of (E-S1, N-S2) and flow dimension of (E-S2, N-S2) [ANOVA; $p < 0.05$]. According to statistically significant results of Table 3, during the first session in the simulator, both tension and negative affect of expert drivers (1.8 and 1.1, respectively) are higher than that of novice drivers (0.6 and 0.6, respectively) while positive affect is lower in expert drivers (1.9) than in novice drivers (3.0). Further, expert drivers have experienced flow more than novice drivers during the second session in the simulator (2.1 and 1.5, respectively). According to the results of other variables, both driver categories have increased their satisfaction of driving in the simulator and their recommendation of using the driving simulator for driver training as well as reporting that their skills have been improved as they practiced more sessions in the simulator.

4.4 Correlation Analysis between Self-Reported Measures of User Experience and Psychophysiological Features

Table 4 contains the significant Pearsons's correlation coefficients between self-reported measures of player experience and psychophysiological measures (P < 0.05). Psychophysiological features represent the seven frequency bands as well as three vigilance states, that is, vigilance high (peaks), vigilance low (valleys), and vigilance difference.

Table 4. Significant correlation coefficients of the correlations between self-reported measures of player experience and psychophysiological measures (P < 0.05). The corresponding EEG frequency bands are D-delta, T-theta, A-alpha, B-beta, and G-gamma.

Variable	Vigilance high	Vigilance low	Vigilance difference
[iGEQ] Competence	(-) B-G [-0.9,-0.95]	(-) D-B-G [-0.88,-0.95]	(-) B2-G [-0.94,-0.97]
[iGEQ] Flow	na	(+) G [0.91]	na
[iGEQ] Tension	(+) B-G [0.91,0.93]	(+) D-B2-G [0.91,0.93]	(+) B-G [0.91,0.92]
[iGEQ] Positive affect	(-) A2-B-G [-0.9,-0.96]	(-) D-B2-G [-0.93,-0.94]	(-) A2-B-G [-0.91,-0.97]
Satisfaction of driving	(-) T-A-B-G [-0.89,-0.93]	(-) T-A-B1 [-0.91,-0.98]	(-) T-A-B-G [-0.89,-0.94]
Whether skills improved?	na	(-) A [-0.91,-0.94]	na

According to Table 4, both flow and tension have significant and positive correlations with psychophysiological features. Conversely, competence, positive affect, satisfaction of driving, and response about the skills report significant but negative correlations with psychophysiological features. However, no significant correlation with any other component was found. The correlations are different based on both polarities of correlations and the frequency bands associated with them.

5 Discussion

The main aim of our study was to compare driving behaviors of expert drivers and novice drivers in a mid-range driving simulator in order to identify measures for evaluating the fitness of driving simulators for driver training. The results in Table 3 were in favor of the aim of our study where both expert and novice drivers have recommended the driving simulator as a training environment in addition to positive reflections about their learning. However, we had to answer two concerns implied by certain results of our study, that is, experts are more adaptive in the simulator environment than novice drivers and/or novice drivers have been over stimulated in the simulator environment to exhibit their mere performance rather than demonstrating their training effectiveness. For instance, Figure 3 showed that novice drivers are attributed with higher degrees of speed, gas, and brake as compared to expert drivers; Figure 4 reveled that expert drivers were more adaptive in the highway traffic track whereas novice drivers were more adaptive in the racing tracks; and Table 3 revealed that although the first driving session was unpleasant to experts, they experienced flow more than novice drivers during their second driving session. Therefore, if we accept the fact that driving behaviors of both driver categories were motivated only by the alternate realism of the simulator instead of the true nature of their driving behaviors, it raises further concerns on the validity of the methodological approach of our study. However, we counter argue the above aspect below.

Although the performance of expert drivers' was much better than novice drivers in the highway traffic track (T30), the order reversed in the racing tracks (T41 and T51). This indicates that experts have exhibited their true skills when the current situation is closer to their real world skillful activity rather than adapting to unfamiliar activities (see [28] for a discussion on expertise). However, according to the results in Table 3, the first driving session seems to have been more awkward to expert drivers than to novice drivers (both tension and negative affect were high) which can be interpreted as experts having more difficulties translating their skills to the new context. The findings are somewhat different with regards to the novice drivers as they have shown competence in their performance in the racing tracks. Moreover, results in Table 3 do not show much negativity, as did the experts in their first driving session or consecutive sessions. Therefore, it seems the novice drivers were intrinsically more adaptive to situations that could be labeled or understood as more fun for them. However, as the results show that their performance was not equal to that of expert drivers in the highway traffic track, we suspect that the highway traffic track came to act as a control to their mere performance oriented motivation while intertwining a learning challenge with it. As the findings related to both categories of drivers exhibit the potential of the highway traffic track to act as an effective training task, we can further assent in favor of driving simulator's capability to enhance learning depending on the characteristics of the task rather than its fidelity.

One of the major limitations of our study is that we could not find more homogenous groups what regards age. This has made us to select a group of teenagers as novice drivers. As a result, some of the aspects such as the dissemination of technology and perception about new digital media (see [37] for a discussion about the difference between digital natives and digital immigrant instructors) were different between the two driver groups. However, we do not see this as a big issue as the situation is very close to the real world. Apart from the above, we found that none of the other conditions such as the sources of disturbances had significant effects on the experimental conditions.

The results in Table 3 and Table 4 were very helpful to verify our data analysis techniques as well as to contextualize the results. For instance, according to the results in Table 3 both tension and negative affect have opposite trends to positive affect, which is indeed the expected pattern of these variables. Moreover, Table 4 confirms the validity of the technique we used to analyze EEG data. For instance, both flow and tension had opposite effects compared to competence and positive affect which is also observable in Table 3. The flow seemed to be associated with focused concentration and learning (gamma band) while tension was identified as a state mediated by both flow and normal state of wakefulness (both gamma and beta bands). The literature to some extent confirms this result because flow makes one to forget about time while being completely absorbed into an activity (see [38]). Moreover, the flow experience in a task is conditioned on certain factors such as a balance between the perceived challenge of a task and the perceived skills of the performer, and, therefore, it needs not to be exhibited in every situation especially in learning situations (see [39]). This relationship confirms the result that expert drivers experienced flow more than novice drivers and that there is a significant difference of how flow and tension have affected both driver groups. Moreover, Table 4 confirms our argument that the recommendations given by the drivers about using the simulator for driver training is unbiased because the results reveals that the variables about satisfaction of driving and reflection about skills improvement have been associated with different activation patterns of the brain than other iGEQ dimensions.

6 Conclusions

In this paper we compared the driving behaviors of expert drivers and novice drivers in a mid-range driving simulator with the intension of evaluating the validity of driving simulators for driver training. The results showed somewhat different driving behaviors of the two categories of drivers, that is, novice drivers were motivated towards performance oriented behavior whereas expert drivers performed well when the task became closer to their regular driving situation. Moreover, we were able to show that the learning effectiveness of simulations heavily depends on the characteristics of simulations rather than their fidelity. Other than the results we presented, we proposed a novel analytical framework for evaluating various aspects of driving behavior.

Acknowledgments. The authors wish to sincerely thank the staff of InGaMe Lab/Interaction Lab at University of Skövde, staff members who voluntarily participated in the experiment, and Fästningens Trafikskola, Karlsborg. This work has been financed through the NeLC project of SPIDER program and internal funding of University of Skövde.

References

1. NHTSA: Traffic Safety Facts, http://www-nrd.nhtsa.dot.gov/Pubs/811856.pdf
2. Collet C., Petit C., Priez A., Dittmar A.: Stroop color-word test, arousal, electrodermal activity and performance in a critical driving situation, Biological Psychology, vol. 69(2), pp. 195--203 (2005)
3. Lal S.K.L., Craig A.: A critical review of the psychophysiology of driver fatigue, Biological Psychology, pp. 173--194 (2001)
4. Bell B.S., Kanar A.M., Kozlowski S.W.J.: Current issues and future directions in simulation-based training in North America. International Journal of Human Resource Management, 19(8), pp. 1416--1434 (2008)
5. Clark C.D.: The Principles of Game Based Learning, in Proc. NETC / LSC Conference, Crystal City, VA, (2004)

6. Backlund P., Engström H., Johannesson M., Lebram M.: Games for traffic education: An experimental study in a game-based driving simulator, Simulation & Gaming, vol. 41(2), pp. 145--169 (2010)
7. SWOV Fact sheet: Simulators in driver training, (2010), http://www.swov.nl/rapport/Factsheets/UK/FS_Simulators_in_driver_training.pdf
8. Backlund P., Engström H., Johannesson M., Lebram M., Sjödén B.: Designing for self-efficacy in a game based simulator: An experimental study and its implications for serious games design, in Proc. Vis'2008, pp. 106--113 (2008)
9. ITS, Program Develops New Test Track Capability, The Sensor Newsletter, Winter (2004)
10. Strayer D.L., Drewa A.: Simulator Training Improves Driver Efficiency: Transfer from the Simulator to the Real World, in Proc. 2nd International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, pp. 190--193 (2003)
11. Palinko O., Kun A.L., Shyrokov A., Heeman P.: Estimating Cognitive Load Using Remote Eye Tracking in a Driving Simulator, in Proc. Eye Tracking Research and Applications, pp. 141--144 (2010)
12. Feinstein A.H., Cannon H.M.: Fidelity, Verifiability, and Validity of Simulation: Constructs for Evaluation, Wayne State University Marketing Department Working Paper 2001-006, (2001)
13. Ritterfeld U., Cody M., Vorderer P.: Serious Games: Explication of an Oxymoron: Introduction, in Serious Games: Mechanics and Effects, Ritterfeld U., Cody M., Vorderer P., eds., New York: Routledge, pp. 3--9 (2009)
14. Ekanayake H.B., Backlund P., Ziemke T., Ramberg R., Hewagamage K.P.: Comparing Expert Driving Behavior in Real World and Simulator Contexts. International Journal of Computer Games Technology, vol. 2013, Article ID 891431 (2013)
15. Hempel C.: Philosophy of Natural Science, Englewood Cliffs, N.J.: Prentice Hall, (1966)
16. Fellows R., Liu, A.: Research methods for construction, Blackwell, UK, (2003)
17. Kolb D.A.: Experiential learning: experience as the source of learning and development. Englewood Cliffs, NJ: Prentice Hall, (1984)
18. Crookall D., Thorngate W.: Acting, knowing, learning, simulating, gaming. Simulation & Gaming, 40(1), pp. 8--26 (2009)
19. Brennecke A.: A General Framework for Digital Game-Based Training Systems, PhD dissertation, Univ. of Rostock, (2009)
20. Thiffault P., Bergeron J.: Monotony of Road Environment and Driver Fatigue: A Simulator Study, Accident Analysis and Prevention, vol. 35, pp. 381--391, (2003)
21. Lin C.T., Wu R.C., Liang S.F., Chao W.H., Chen Y.J., Jung T.P.: EEG-based drowsiness estimation for safety driving using independent component analysis, in Proc. IEEE Trans. Circuits Syst. I, vol. 52(12), pp. 2726--2738, December (2005)
22. Gruyer D., Rakotonirainy A., Vrignon J.: Data combination using Belief Theory to Assess Driver's Vigilance, in Proc. Australasian Road Safety Research, Policing and Education, Wellington, New Zealand, November (2005)
23. Roge J., Pebayle T., Lambilliotte E., Spitzenstetter F., Giselbrecht D., Muzet A.: Influence of Age, Speed and Duration of Monotonous Driving Task in Traffic on the Driver's Useful Visual Field, Vision Research, vol. 44(23), pp. 2737--2744 (2004)
24. Egen T., Lervåg L., Moen T.: Evaluation of IVIS/ADAS Using Driving Simulators: Comparing performance measures in different environments, in Proc. European Transport Conference, Netherlands, (2009)
25. Fairclough S.H.: Psychophysiological Inference and Physiological Computer Games, in Proc. Brain-Computer Interfaces and Games (Brainplay '07), Salzburg, Austria, (2007)
26. Cacioppo J.T., Tassinary L.G., Berntson G.G.: Handbook of psychophysiology, 3rd ed., New York: Cambridge University Press, pp. 59--61 (2007)

27. Picard R., Papert S., Bender W., Blumberg B., Breazeal C., Cavallo D., Machover T., Resnick M., Roy D., Strohecker C.: Affective learning: A manifesto, BT Technology Journal, vol. 22(4), pp. 253--264 (2004)
28. Dreyfus H.L.: A Phenomenology of Skill Acquisition as the basis for a Merleau-Pontian Non-representationalist Cognitive Science, in Proc. Foundations and the Ontological Quest., Rome, Vatican City, (2002)
29. Hudlicka, E.: Affective Game Engines: Motivation and Requirements, in Proc. 4th International Conference on Foundations of Digital Games, Orlando, Florida, pp. 299--306, (2009)
30. MATLAB version 7.6.0. Natick, Massachusetts: The MathWorks Inc., (2003)
31. Delorme A., Fernsler T., Serby H., Makeig S.: EEGLAB Tutorial, Univ. of San Diego California, (2006)
32. Bos D.O.: EEG-based emotion recognition, The influence of Visual and Auditory Stimuli, (2006)
33. Shi L.C., Lu B.L.: Off-Line and On-Line Vigilance Estimation Based on Linear Dynamical System and Manifold Learning, in Proc. 32nd International Conference of the IEEE Engineering in Medicine and Biology Society, Buenos Aires, Argentina, pp. 6587--6590, (2010)
34. Nie D., Wang X.W., Shi L.C., Lu B.L.: EEG-based Emotion Recognition during Watching Movies, in Proc. 5th International IEEE/EMBS Conference on Neural Engineering, pp. 186--191, (2011)
35. IJsselsteijn W., Hoogen W.V.D., Klimmt C., Kort Y.D., Lindley C., Mathiak K., Poels K., Ravaja N., Turpeinen M., Vorderer P.: Measuring the Experience of Digital Game Enjoyment, in Proc. Measuring Behavior, Maastricht, The Netherland, August (2008)
36. Drachen A., Nacke L.E., Yannakakis G., Pedersen A.L.: Correlation between heart rate, electrodermal activity and player experience in First-Person Shooter games, ACM SIGGRAPH Video Game Symposium, Los Angeles, pp. 49--54 (2010)
37. Stricker A.G.: Why Affective Learning in a Situated Place Matters for the Millennial Generation, (2009), http://www.au.af.mil/au/awc/awcgate/a46/affective-learning-situated-place.pdf
38. Geirland J.: Go with the flow, Wired, vol. 4(9), pp. 160--161 (1966), http://www.wired.com/wired/archive/4.09/czik_pr.html
39. Kort B., Reilly R., Picard R.W.: An affective model of interplay between emotions and learning: Reengineering educational pedagogy-building a learning companion, in Proc. Advanced Learning Technologies, pp. 43--46. IEEE Computer Society (2001)
40. McDonald J.H.: Handbook of Biological Statistics, 2nd ed., Baltimore: Sparky House Publishing, (2009)