BACHELORTHESIS, AUGUST 2012

EXPLORING ELECTRODERMAL ACTIVIY, EXPERIENCED WORKLOAD AND PERFORMANCE DURING SIMULATOR TRAINING

Martijn Broeks (s0175315)

FACULTY OF BEHAVIORAL SCIENCES

DEPARTMENT COGNITIVE PSYCHOLOGY AND ERGONOMICS

EXAMINATION COMMITTEE

First supervisor: Dr. Matthijs Noordzij

Second supervisor: Teun Lucassen, Msc

University of Twente

Abstract

Novice drivers start their driving experience at a driving school. They learn their skills in a car under the supervision of a driving instructor. In recent years a company called Green Dino developed a driving simulator to teach novice drivers the basic skills. The driving simulator has proven itself as a tool to increase the change to pass the drivers exam. The aim of this study is to observe novice drivers during their driving simulator course. Three male and two female drivers, age 17 and 18, participated in the study. The EDA levels were acquired using an *Affectiva* Q-sensor, which is an unobtrusive wristband with a wireless biosensor. These EDA levels are a physiological way to measure cognitive demand or arousal and are also used to look at habituation effects. The NASA-TLX test provided a subjective way to measure cognitive demand. A performance rating came from the simulator that grades student performance. Three simulator variables are also analyzed, the lane position, the distance to a preceding vehicle and the distance to a new road element. The results partly agree with the literature. The participants differ in their EDA levels and habituation effects are found as participants show a drop in EDA levels over time. The performance compared with the EDA levels show a result that is consent with the Yerkes-Dodson law, an inverse u-shaped relation between performance and arousal. Against expectations there was no relation found between the NASA-TLX subscales and the EDA levels. Two of the three simulator variables relate significantly with the EDA levels, the lane position and the distance to a new road element. At an individual level the relations with the EDA levels differ. Particpant 3's EDA levels have a significant relation with all the simulator variables, whereas the EDA levels from participant 4 show no significant relation with the simulator variables. The overall results of this study indicate that EDA measurement by means of an *Affectiva* Q-sensor is a useful and valid approach to assess EDA levels of novice drivers during a driving simulator course. Furthermore, this study shows that a driving simulator causes EDA responses in participants which relate to performance ratings, as novice drivers respond to specific driving situations. This explorative study reveals new possibilities for further research and is relevant for the simulator industry, driving schools, traffic psychology and psychophysiology research.

Samenvatting

Beginnende bestuurders krijgen hun eerste rijervaring bij een rijschool. Hier leren ze de vaardigheden om een auto te besturen onderbegeleiding van een rijinstructeur. In de afgelopen jaren heeft Green Dino een rijsimulator ontwikkeld om beginnende bestuurders de basis vaardigheden te leren. De kans om te slagen stijgt door het gebruik van de rijsimulator. Het doel van deze studie is om beginnende bestuurders te observeren gedurende hun rijsimulator cursus. Drie mannelijke en twee vrouwelijke bestuurders van 17 en 18 jaar oud nemen deel aan deze studie. EDA levels zijn gemeten met een Q-sensor van *Affectiva*. Dit is een nauwelijks merkbare armband met draadloze biosensoren. Het meten van EDA is een fysiologische manier om cognitieve belasting of arousal te meten. Daarnaast kan met de EDA levels gekeken worden naar habituatie effecten. Cognitieve belasting wordt ook gemeten door middel van een subjectieve NASA-TLX test. De simulator produceert ook een performance score die de rijvaardigheid van de student beoordeelt. Drie variabelen uit de simulator worden ook meegenomen, namelijk: Positie binnen de rijbaan, afstand tot het voorgaande voortuig en de afstand tot een nieuw weg type. De resultaten van het onderzoek komen gedeeltelijk overeen met de gevonden literatuur. Proefpersonen verschillen van elkaar in hun EDA levels en vertonen ook habituatie. Proefpersonen hebben een daling in EDA levels naarmate de tijd vordert. Wanneer de prestaties worden vergeleken met de EDA levels is het resultaat consistent met de Yerkes-Dodson Law, die stelt dat de relatie tussen prestaties en arousal gelijk aan een geinverteerde U. Tegen de verwachtingen in is er geen relatie gevonden tussen de NASA-TLX subschalen en de EDA levels. Twee van de drie simulator variabelen hebben wel een significante relatie met de EDA levels, namelijk de positie op de weg en de afstand tot het volgende weg type. Op een individueel niveau zijn er grote verschillen tussen de proefpersonen. De EDA levels van Proefpersoon 3 vertonen allemaal een significante relatie met de simulator variabelen. Bij proefpersoon 4 daarin tegen hebben de EDA levels geen significante relatie met de simulator variabelen. Deze studie laat zien dat de EDA metingen van de *Affectiva* Q-sensor nuttige zijn en valide informatie opleveren wanneer het gaat om het bepalen van EDA levels gedurende een rijsimulator cursus. Verder laat deze studie zien dat een rijsimulator voor EDA response zorgt bij proefpersonen. Deze response is weer te relateren aan prestaties en wanneer beginnende bestuurders een andere weg situatie tegenkomen. Deze exploratieve studie laat nieuwe mogelijkheden zien voor vervolg onderzoek en is relevant in de simulator industrie, bij rijscholen, verkeerspsychologie en psychofysiologisch onderzoek.

Table of Contents

1. Introduction

1.1 Novice drivers

In the Netherlands young adults are able to take their drivers exam at the age of 17. However due to high expectations of novice drivers many young adults start taking driving lessons before the official exam. The goal of these lessons is to teach novice drivers the skills which enable them to drive a car and to navigate safely through traffic. The "Centraal Bureau Rijvaardigheidsbewijzen" (CBR) is responsible for reviewing the driver exams and to grant them their driver's license. In 2011 the CBR took 487.685 car driver's exams of which 217.611 passed. This means 44.6% got their driver's license (Hart & Staveland, 1988). However, in previous years it has been found that novice drivers who have taken simulator drivers lessons in advance, have a 4-5% higher chance of passing the drivers exam (Howarth & Hodder, 2008).

The present study will focus on the simulation lessons for novice drivers as well, by observing them during their simulation course. During the observations, the arousal levels, experienced workload and performance have been measured. The aim of this study is to look for relations between the arousal levels, experienced workload and performance levels. There are other studies which have researched these variables (e.g., Katzourakis, Abbink, Happee, & Holweg, 2011; Mantovani, 2001; Nagai, Critchley, Featherstone, Trimble, & Dolan, 2004). Unlike other studies, this study also takes changes over time into consideration. This extra variable is unique since arousal levels are monitored over time in a real life situation. This study can be considered as an explorative ideographical study. The following introduction will explain the concepts and background of this study

1.2 Driving simulator

Virtual reality is a concept that can be used for the training of for instance practical skills (Mantovani, 2001). People are able to learn how to perform certain tasks in a virtual reality environment. The acquired knowledge and skills have shown to transfer to performances in the real world as well (Psotka, 1996). Real world training and virtual environment training have shown equivalent levels of performance, while the performance of people with no training at all, was significantly lower (Rose, Attree, Brooks, Parslow, & Penn, 2000). Therefore, virtual reality represents an area with high potential of enhancing and modifying the learning experience. It allows entry to entirely new capabilities and experiences, which would be too difficult, too costly or simply impossible to enhance in the real world. Virtual reality environments can be tailored to individual learning- and performance styles. They are highly flexible and programmable, enabling the teacher to present a wide variety of controlled stimuli and to measure and monitor a wide variety of responses produced by the user. Virtual reality environments can provide a rich, interactive, engaging educational context, supporting experiential learning (Mantovani, 2001). This concept is used in traffic research and education in the form of driving simulators. Driving simulators are widely used for applications such as driver behavior and perception research, vehicle development, steering systems prototyping, and drivers training (De Groot, De Winter, García, Mulder, & Wieringa, 2011; Hoedemaeker & Brookhuis, 1998; Katzourakis, et al., 2011; Kemeny & Panerai, 2003).

Driving simulators are virtual reality tools, which gives the driver on board the impression that he/she is driving an actual vehicle by predicting vehicle motion caused by driver input and feeding back corresponding visual, motion, audio and proprioceptive cues to the driver (W. S. Lee, Kim, & Cho, 1998). Simulators offer features that are not easily attainable through on-the-road training, such as objective performance measurement, manipulation of the environment according to the learning goals, and exposure to errors and hazardous driving conditions in a controlled and repetitive manner without physical risk (Allen, Park, Cook, & Fiorentino, 2007). Although there are no real dangers in a simulator and errors do not lead to disadvantageous consequences, people do react to the diverse simulator situations with different levels of arousal (Kemeny & Panerai, 2003).

1.3 Arousal

The level of arousal describes the global state of the brain. The broadest categorization is whether an individual is awake or asleep. However an individual that is awake can be operating at different levels of arousal (Picard, 2009). A certain level of arousal is required for an individual to perform fully. This is described in the Yerkes-Dodson law. This law states that Performance increases with arousal when the arousal level is low, then reaches its peak at the optimal arousal level, and then decreases as arousal continues to increase, as seen in Figure 1 (Yerkes & Dodson, 1908).

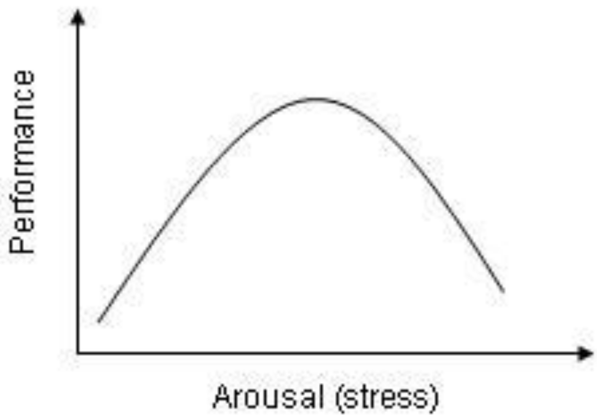


Figure 1. The Yerkes-Dodson Law

Another study describes arousal as a physiological or psychological state of being excited or activated (Vaez Mousavi, Naji, Osanlou, & Esmaeilpour Marandi, 2011). Physiological arousal is characterized by heightened activation of the autonomic nervous system, which generally implies the degree of physiologically evoked arousal response to stimuli that is determined by measuring physiological variables such as skin conductance, muscle movement, or the heart rate (Wrisberg, 1994).

1.4 Electrodermal Activity (EDA)

In the present study skin conductance will be used as a measurement for arousal. Skin conductance can be used to show arousal in a virtual environment (Jang et al., 2002). A more common term for skin conductance is EDA. EDA describes all electrical phenomena in the skin (Johnson & Lubin, 1966). One of these methods involves two electrodes that are placed on the skin's surface, which wwill send a small electrical signal through the skin. The conductivity of the skin is expressed in skin conductance and is measured in small time intervals. There are two EDA phenomena, tonic and phasic, which differ in skin conductance variation. A slowly varying tonic activity (skin conductance level; SCL) reflects the overall degree of arousal that decreases with physiological relaxation such as rest and sleep. A fast varying phasic activity (Skin conductance response; SCR) is widely used in experimental psychological studies as an objective measure of conscious and unconscious emotional processing and attention. The term response suggests that there is a stimulus preceding it. A SCR that appears up to 5 seconds after the beginning of an intentional of unintentional stimulus can be considered a SCR. However, it is often the case that a SCR cannot be reduced to one specific stimulus. That is why these SCR's are called nonspecific skin conductance responses (NS-SCR) (Boucsein, 2011).

1.5 Habituation

Interactive virtual environments are known to cause symptoms similar to those of motion sickness (Jang, et al., 2002). A repeated exposure to virtual environments causes a decrease in the level of symptoms (Johnson & Lubin, 1966). The general process of a decrease in symptoms or reactions to the same stimuli (for instance caused by a virtual environment) after repeated exposure is also referred to as habituation (Nagai, et al., 2004).

The foundation of habituation can be found in the orienting response. The orienting response is an autonomic response to new stimulus and indirectly reflects how much a person attends to and processes new stimulus in the environment (Cacioppo, Tassinary, & Berntson, 2007). Certain features of a new stimulus are stored in neuronal networks in the brain, creating a "neuronal model". This information is also forwarded to novelty detectors in the hippocampus. When another stimulus is presented it will be compared with the excising neuronal models. When a mismatch occurs the stimulus will be processed like any other new stimuli. When the same stimulus is presented repeatedly, it will continue to be analyzed. However, this will cause less and less activity in the novelty detectors in the hippocampus (Boucsein, 2011). In comparison with other physiological indicators of an orienting response, the phasic EDA component is the most suitable. The phasic EDA component is the indicator of habituation most frequently used (Boucsein, 2011). One of the characteristics of habituation is the response decrement to a repeated stimulus. Habituation to a stimulus occurs with the decrease in SCR (Steiner & Barry, 2011).

1.6 Workload

The term workload represents the cost of accomplishing a task for an individual (Hart & Staveland, 1988). It is determined by the interaction of task demands, the conditions under which it is performed and skills, behaviors and perceptions of the individual (Malmo, 1959). Defining workload as a demand that is placed on an operator is somewhat simplistic in that it attributes workload exclusively to an external source (Cacioppo, et al., 2007). Demands of a task may include physical actions and/or cognitive tasks. In turn, the impact of these demands is dependent on the abilities of an individual performing the task (Malmo, 1959). Several studies have shown a relation between EDA and demand in simulators. Research in flight simulators have shown that EDA correlates highly with heart rate. Furthermore, the both of them exhibited changes in to the various demands of the flights. The high correlation between the two factors is due to a common underlying nervous system (Kemeny & Panerai, 2003). A summary of the study from the Human machine interface And the Safety of Traffic in Europe (HASTE) project have shown the impact of visual and secondary cognitive tasks on performance and physiology in both simulated and on-road driving. In the simulator, skin conductance was found to be increasing during a visual-based dual task (Hoedemaeker & Brookhuis, 1998). EDA varies with the changing driving demands. With higher demands from a driving task comes an increase in the skin conductance level. EDA remained unchanged or even decreased when cognitive demands were at the lowest level (De Groot, et al., 2011; Katzourakis, et al., 2011). This was initially found in young drivers, but applies to ages differing from 20 to 69 (Meshkati, Hancock, Rahimi, & Dawes, 1995).

1.7 Subjective measures of workload

The EDA measurements are a physiological way to measure workload. However, there are other ways to measure workload as well. The subjective measure is one of them and assumes that workload is linked to perceived effort and can be assessed by individuals (Meshkati, et al., 1995). Subjective measures are becoming increasingly important in system evaluations and have been used extensively to assess operator workload. The reason subjective measures are frequently used is because of their practical advantages (ease of implementation and non-intrusiveness) and their capability to provide sensitive measures of operator workload are different procedures to assess mental workload.

Many subjective procedures exist to measure mental workload. The most outstanding are the Cooper-Harper Scale, the Bedford Scale, The SWAT(subjective Assessment Technique) and the NASA-TLX(Task Load Index) (Rubio, Díaz, Martín, & Puente, 2004). The NASA-TLX is chosen to be used for this research. The NASA-TLX is a multi-dimensional scale designed to obtain workload scores estimates from one or more operators immediately after a task. The NASA-TLX test has proven itself in various study areas as a valid test to assess workload (Hart & Staveland, 1988).

1.8 This research

The goal of this explorative study is to observe novice drivers during their driving simulator course. The novice drivers agreed to be observed but there main goal remains to acquire skills and insight. This creates an unique setting where novice drivers can be observed during their learning process. The novice drivers have been asked to participate in this research, so there are no selection bias influences.

The EDA data, NASA-TLX and performance scores provide different insights in the behavior of the participants in the simulator. EDA data will be analyzed for habituation effects and individual differences. The different measurements will also be compared with each other. The first hypothesis of this research is aimed at the EDA data and the NASA TLX as these two variables both measure cognitive workload although in a different way. Therefore, one would expect a relation between the two variables. The NASA-TLX uses six subscales, of which two (Mental Demand and Frustration) are selected to compare with the EDA data. Because these variables where expected to have the best correlation with electrodermal activity. Mental demand is the scale on how much mental and perceptual activity is required and frustration level is the scale on how insecure, discouraged, irritated, stressed and annoyed one feels. Therefore the first hypothesis is: “Is there a relation between the EDA levels and the NASA-TLX variables, ‘Mental Demand’ and ‘Frustration’?”

The simulator output contains information on various movements inside the virtual environment. The other hypotheses are aimed at seeing whether the movements inside the virtual environment have any effect on the cognitive demand. These simulator variables will be explained in the method section.

Cognitive load is known to influence overall driving performance (Mehler, Reimer, Coughlin, & Dusek, 2009). Various driving task show a relation with cognitive demand

(Schießl, Vollrath, Altmüller, Dambier, & Kornblum, 2006). In some studies, cognitive distraction has increased lane variability. For instance, a sentence comprehension task increased drivers' lane keeping errors as compared to performance during single task driving (Just, Keller, & Cynkar, 2008). A word generation task increased the amount of errors in a pursuit tracking task (Strayer & Johnston, 2001). Surprisingly , the opposite effect was found in other studies. Cognitive distraction decreased the lane position variation. A challenging working memory task improved lane maintenance by decreasing standard deviation of lane position (Brookhuis, de Vries, & de Waard, 1991). Drivers deviated less from their lane center and committed fewer lane violations (crossing center-line or road edge) when engaged in either a cell phone conversation or a signal detection task than when driving under single-task conditions (Beede & Kass, 2006). Similar effects of cognitive distraction on lane keeping have been found, when using an *n*-back counting task or an auditory continuous memory task (Engström, Johansson, & Östlund, 2005; Jamson & Merat, 2005; Reimer, 2009). These contradictions in the study results make the lane variance an interesting variable. However, in this driving simulator course the participants will be subjected to different distracters. So the cognitive demand is determined by the overall level of arousal. This lead to the following hypothesis: "Is there a relation between the EDA levels and the position of the participant on the road?"

A previous study found that following a car causes higher cognitive demand, depending on the type of road. Regulating a safe distance with the preceding car did not cause higher cognitive demand (Schießl, et al., 2006). The third hypothesis focuses on cognitive load and the preceding vehicle. Cognitive demand is again determined by the overall level of arousal. Leading to the following hypothesis: "Is there a relation between the EDA levels and the distance to the next vehicle?"

The road environment affects the cognitive load experienced by drivers (Schießl, et al., 2006). The cognitive demand differs between different road types (De Waard, Jessurun, Steyvers, Raggatt, & Brookhuis, 1995). Complex road environments are more demanding than simple road environments (Horberry, Anderson, Regan, Triggs, & Brown, 2006). The last hypothesis looks at the cognitive demand leading up to a new type of road. These new road types can vary from a rural road to an intersection. The effect of the different road types is not analyzed. This leads to the following hypothesis: "Is there a relation between the EDA levels and the distance to the next road element?"

2. Method

2.1 Participants

The participants in this research are young adults who will start with their driving course. Participants are recruited among students that started at the driving school ”Auto en Motor Rijschool Lucassen”. The driving school offers regular driving lessons and simulator driving lessons. Students who have started with the simulator driving lessons were asked to participate in the research. A maximum of five participants could participate in the study. All participants have no previous driving experience at another driving school. Three male and two female students participate in the research, four of them are 17 years old and one of them is 18 years of age. They are required to read and sign an informed consent form. If the participant is under the age of 18, a signature from one of the parents is required.

2.2 Materials

The research is conducted using a driving simulator constructed by Green Dino. The driving simulator “Classic” (see Figure 2) consist of a metal framework with three beamers projecting a 180° virtual environment in front of the driver (GreenDino, 2012). The vehicle controls of the simulator resemble that of a real car with a manual gearbox. Force feedback was provided on the steering wheel and acceleration cue were supplied by vibrating elements in the steering wheel and seat. The dashboard, interior and mirrors were integrated in the projected image. The simulator driving program offers twenty session taking about twenty-five minutes each. The sessions are divided into individual sections that vary from three to twelve minutes. Each session then consists of 3 to 6 sections and is graded at the end of the section. An overall report is available at the end of each session. The simulator provides real time instructions to the participant during the sessions, depending on their driving skills. The simulator records all the actions done by the participant in the driving simulator, together with movements inside the virtual environment. These actions are logged with a sampling frequency of 8Hz. The simulator creates a new log file for each section that is completed by a participant. A webcam recorded the participants inside the simulator.

Figure 2. Example of the driving simulator “classic”

The electrodermal activity of the participants was measured using a Q-sensor. Researchers at MIT developed a wrist worn EDA-sensor that can be used outside laboratory settings (Poh, Swenson, & Picard, 2010). This product is brought to the consumer market by Affectiva in the form of a Q-sensor. The Q-sensor is the size of a watch and has no wires connected to it. Electrodermal activity is measured with two electrodes pressed to the skin. The skin conductance is expressed in micro-Siemens (µS). The Q-sensor also measures skin temperature and acceleration on the X, Y and Z axes. These measurements are done at 32Hz. Using the button on the Q-sensor it is possible to place makers in the EDA-data. To get accurate measurements the Q-sensor has to be worn ten minutes prior to the experiment. The skin between the two electrodes needs to build up an amount of sweat to make resistance measurements feasible.

For the workload assessment a NASA-TLX test was used. The NASA TLX uses six dimensions to assess mental workload: mental demand, physical demand, temporal demand, performance, effort, and frustration. Twenty step bipolar scales are used to obtain ratings for these dimensions. A score from 0 to 100 (assigned to the nearest point 5) is obtained on each scale. A weighting procedure is used to combine the six individual scale ratings into a global score; this procedure requires a paired comparison task to be performed prior to the workload assessments. Paired comparisons require the operator to choose which dimension is more relevant to workload across all pairs of the six dimensions. The number of times a dimension is chosen as more relevant is the weighting of that dimension scale for a given task for that operator. A workload score from 0 to 100 is obtained for each rated task by multiplying the weight by the individual dimension scale score, summing across scales, and dividing by 15 (the total number of paired comparisons).

2.3 Procedure

The research is conducted at the office of ”Auto en Motor Rijschool Lucassen”. During a simulator driving course students follow the lessons at this location. The driving school offers one hour lessons that consist of two sessions. The simulator driving course consists of ten lessons. Students that follow the simulator driving course can take only one lesson each day. The time between the first and the last lesson depends on the number of lessons a student takes in a week. Participant 5 completed the course in only four weeks, Participants 1, 2 and 4 took eight weeks to complete the course and participant 3 took the longest time with 10 weeks.

Students that participate in the research are expected ten minutes prior to their simulator driving lesson. Right upon arrival the Q-sensor was placed, allowing some sweat to build up between the electrodes and the skin. During the ten minutes waiting time the participant read the instruction that explains the upcoming sessions. After ten minutes the participant is seated in the simulator and the video recording starts. Before the start of the lesson the participant is asked to push the button on the Q-sensor to place a marker in the EDA-data. During the sessions the participant is driving the simulator alone. Further instructions will only be given when a participant needs help or when the driving instructor thinks it is needed. At the end of each session the driving instructor will review the scores with the participant. When the participant has completed two sessions the video recording is stopped. After reviewing the last scores with the driving instructor the participant is asked to fill in the NASA-TLX test. After the test the Q-sensor is removed and the participant is done for the day. The data from the simulator, video recording and Q-sensor is stored at a central location. These steps will be repeated for each participant until they have completed all twenty session. During the last session of each participant the driving instructor changed the type of simulator vehicle from a car into a lorry or bus, to give participants an impression of how this vehicle handles.

2.4 Data Analysis

For unknown reasons the simulator was not always recording during a session. This means that for some participants data was missing. A total of four sessions are missing in the dataset. The last session that is done in a lorry or bus is excluded from the dataset. The data that is collected in this session is not relevant for the research.

The remaining data creates a large dataset. This contains data from the simulator, EDA measurements, video images and the NASA-TLX scores. These different resources need to be synchronized. The simulator collects data on position, movements of the car and the position of the driving instruments. It also scores the performance of participants on the driving tasks that were presented. All these variables are logged and can be accessed. The EDA measurement data contains information on the skin conductance, skin temperature, acceleration on the X, Y and Z axle and the time. Also the participant has created markers in each data file. The video images provide an image of the participant in the simulator and also show part of the simulator screen. The NASA-TLX data is collected at the end of each session and is an evaluation by the participant. Data is collected on 5 scales and weighted by their significance according to the participant.

Data from the Q-sensor is converted to text files, this made the data usable in Matlab. Skin conductance is a physiological signal with relatively slow changing levels and sudden peaks associated with discrete arousal events (Reimer, Mehler, Coughlin, Godfrey, & Tan, 2009). The number of peaks is represented in the NS-SCR. A script in Matlab is used to convert the skin conductance units into usable NS-SCR's per minute. This is done by filtering the data with a Butterworth filter (cut-off frequency= 1/32 Hz). Next, the first derivative was taken from this signal. In order to label each NS-SCR the first derivative was searched for data points with a minimum speed difference of .0004 μS/ms. A minimum of 700 ms has to be between each NS-SCR otherwise they would be put together and the peak value has to be bigger than .01 μS.

The simulator collects a large amount of data but due to limited time resources only a few of these variables will be analyzed. The simulator log file data that has been used for this research contains information on the location of the vehicle relative to the center of the driving lane, distance to the preceding vehicle and distance to the next element on the road. The location of the vehicle relative to the center of the driving lane is represented in a strip index. This value can vary from 0 to 3 depending on the number of lanes available. Each round number relates to the center of that specific lane. The distance to the preceding vehicle is a value that show whether or not there is a vehicle in front of the participant and the distance to that vehicle. The distance to the preceding vehicle is expressed in meters and when there is no car in front of the participant this is show with a -1 value. Distance to the next element show the distance from the participant to the next road element type and is expressed in meters. Differences between element types will not be analyzed in this research.

The NASA-TLX has six scales which measure different aspects of the perceived workload. For this research two scales are selected. The scales mental demand and the frustration level are selected for further analyzing.

The video images are used to synchronize the EDA data to the log files. The video images contain information on when a marker is created in the EDA data. It also shows when the simulator starts a new section of the driving course. It is possible through video analysis to determine when a new section of the simulator session starts relative to the EDA marker. This process has been done by two researchers. The null hypothesis in which the two researchers significantly differ has been tested with a *Kruskal Wallis* test. The interobserver reliability showed no significant differences between the two researchers, $\kappa = .787$,$p < .01$. The synchronized data can be used for analyze in Matlab. The NASA-TLX file contains information on one lesson. This means that the EDA data of two sessions is synchronized to each NASA-TLX test.

Statistical comparison has been computed with SPSS 18 using a *Kolmogrov-Smirnov* test to see if normality can be assumed. Further analysis is done using the *Kruskal Wallis* test, linear regression analysis, quadratic regression analysis, graph plotting and comparison of mean ranks. Microsoft Excel has been used to display the EDA data in a histogram and a comparison of mean ranks is done on the EDA data. The EDA parameters are plotted over time. Habituation effects are determent with a qualitative analysis, by determining whether a drop in the physiological response occurs.

Performance and the EDA data were compared using a linear regression analysis. The data was also plotted with driving performance dependent on the EDA data. The performance was also compared with the NASA-TLX scores. Another linear regression analysis was carried out on the performance with the NASA-TLX scores. The NASA-TLX was also compared with the EDA data with a linear regression. Three simulator variables were compared with the EDA data with another linear regression.

3. Results

Before carrying out any statistical analyzes, the assumption of normality was tested. This was done using a *Kolmogrov-Smirnov* test. Both EDA variables were tested, number of NS-SCR's Z= 10.01,p < .001 and the total amplitude of NS-SCR's Z= 15.14,p < .001. The hypothesis that normality can be assumed was rejected. With no normality assumed the data will be analyzed with non-parametric tests.

3.1 EDA

The electrodermal activity differs from person to person. This research contains the EDA data of 5 participants. To check if there are any differences between de participants a Kruskal Wallis test is preformed on the data. The null hypotheses is rejected, meaning that there are significant differences between the participants, χ^2 =528.37,p <.001. These differences are clearly show in Figure 3 and Figure 4

Figure 3. Overview of the total quantity of NS-SCR's per lesson (each bar represents the total number of NS-SCR's for that lesson) and per participant

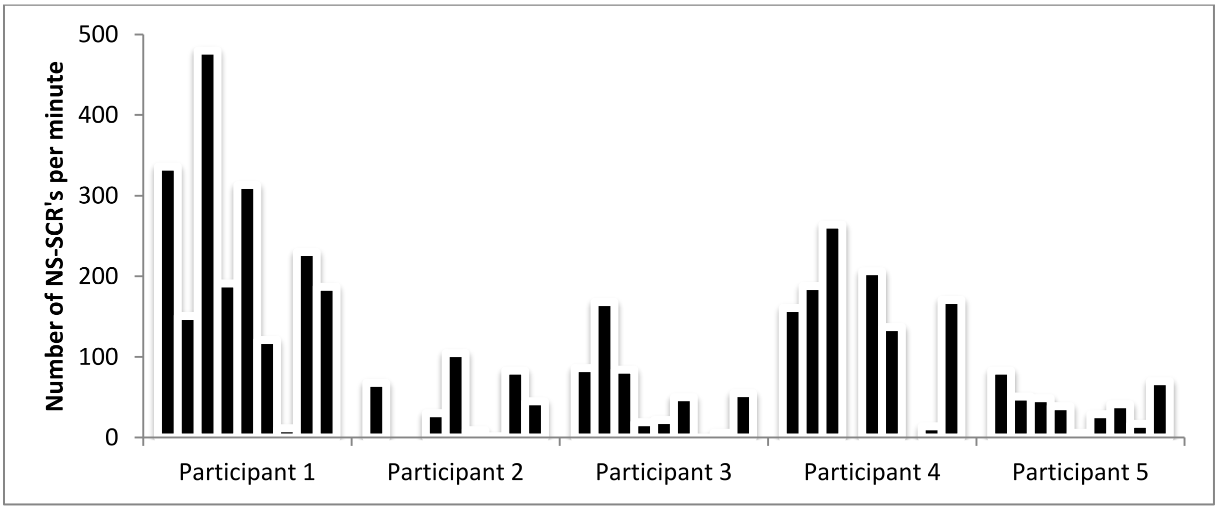
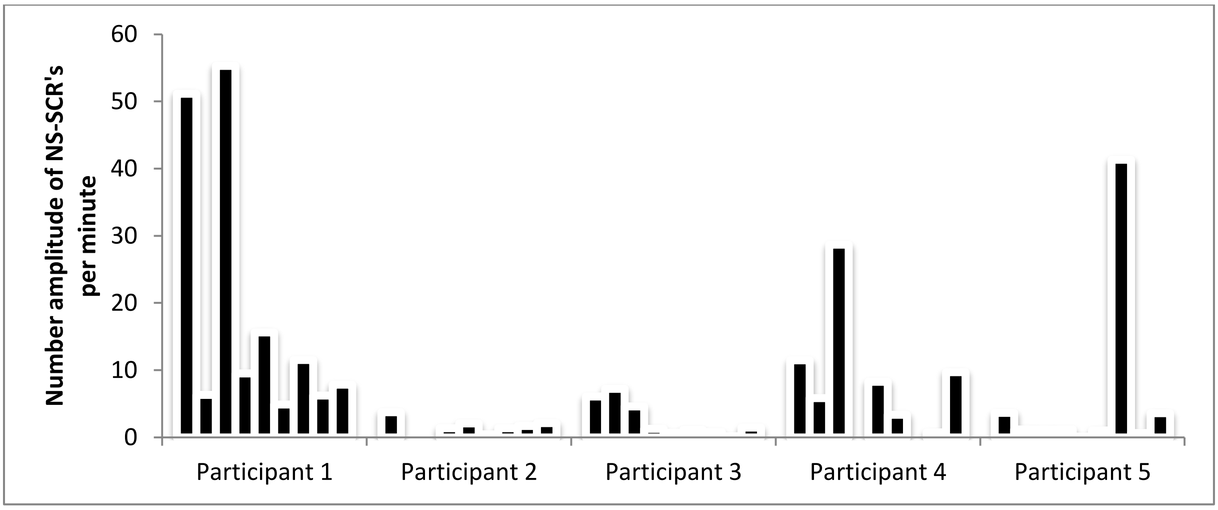


Figure 4. Overview of the total amplitude of NS-SCR's per lesson (each bar represents the total amplitude of NS-SCR's for that lesson) and per participant



Further descriptive statistics were applied on the EDA-data. The data consist of lesson 1 to 9. The highest number of NS-SCR during a lesson was 480 NS-SCR's by participant 1, followed by participant 5 with 342 NS-SCR's and participant 4 with 264 NS-SCR's. Lowest number of NS-SCR's belonging to participant 3 with 168 NS-SCR's and participant 2 with 105 NS-SCR's. The ranking is the same for the amplitude of the NS-SCR's. Participant 1 has an amplitude 55.29 NS-SCR's, participant 5 has an amplitude 41.32 NS-SCR's, participant 4 has an amplitude 28.68 NS-SCR's, participant 3 has an amplitude 7.25 NS-SCR's and participant 2 has an amplitude 3.78 NS-SCR's. An overall descriptive statistics is show in table 1.

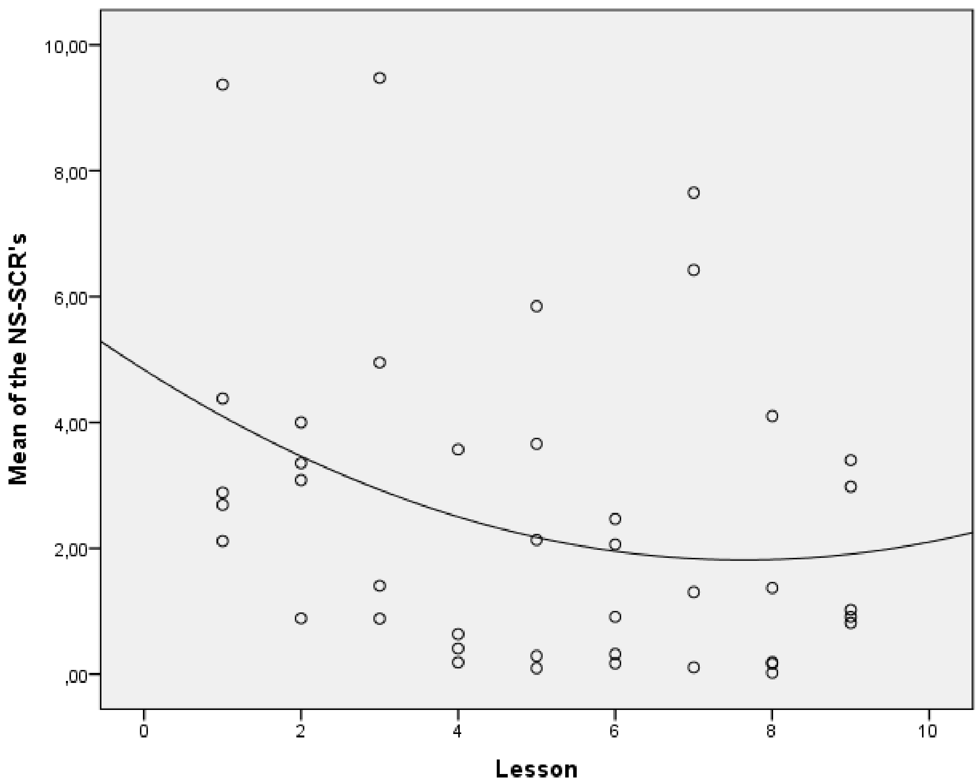
Table 1. Descriptive statistics for the total of NS-SCR's and the total amplitude of NS-SCR's

Variable	N	M	Mdn	S.E of Mean	Min	Max
NS-SCR's quantity	41	126	83	18	6	480
Total amplitude of NS-SCR's	41	7,98	3,62	2,03	,16	55,29

Participants differ from each other in their electrodermal activity. But each individual can also differ in their electrodermal activity between lessons. Due to a habituation effect it is possible that participants had a decrease in electrodermal activity. A significant linear regression was found for both between lessons and during lessons. However a R^2 of .02 was found for both the total of NS-SCR’s per minute and the amplitude of NS-SCR’s per minute.

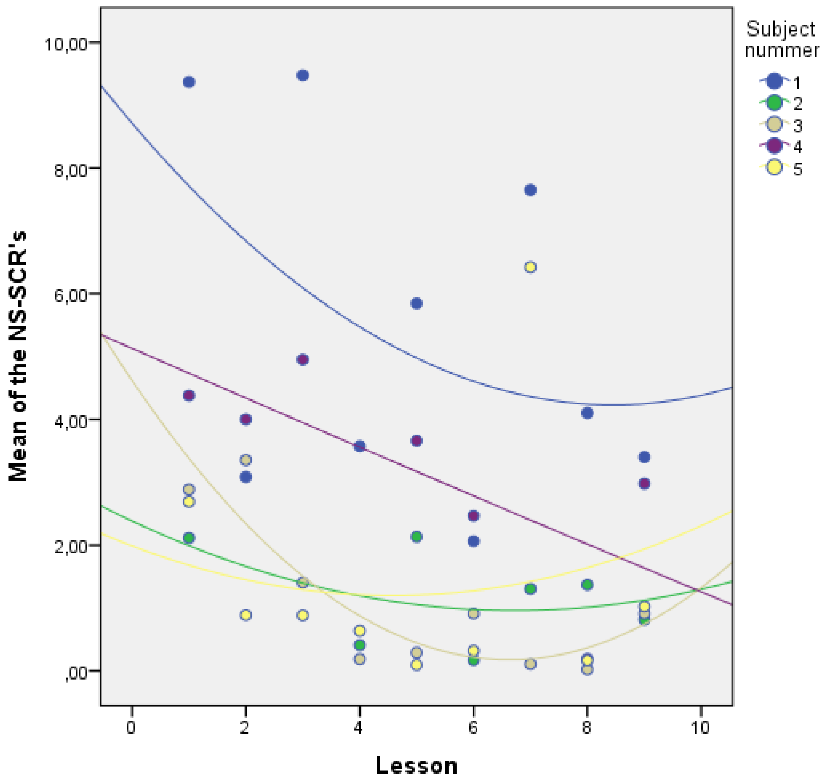
In some analysis the total of NS-SCR’s per lesson is not representative. Because each lesson differs in length, so a difference in total of NS-SCR’s can also be caused by a difference in length. For these analyses an average NS-SCR per lesson is a good measure to compare lessons and individuals. Figure 5 illustrates the mean of the NS-SCR per lesson. The figure shows a drop in means of NS-SCR for the participants over time. This is done with a quadratic regression analysis.

Figure 5. Mean of the NS-SCR per minute of each participant for each lesson (with quadratic regression line).



Because of the individual differences a quadratic regression analyses is also done for each participant. In figure 6 the mean of the NS-SCR per lesson are marked for each participant and a quadratic regression analysis is done for each participant. And looking at individual differences, it’s clear to see that some participants show a larger decrease in the mean of NS-SCR’s than others. Participants 1and 3 have the largest decrease in the mean NS-SCR, whereas participant 2 and 5 show little to no decrease in the mean NS-SCR. And although this analysis is done with a quadratic regression analysis the regression line of participant 4 appears to be linear.

Figure 6. Mean of the NS-SCR per minute of each participant for each lesson (with quadratic regression line).

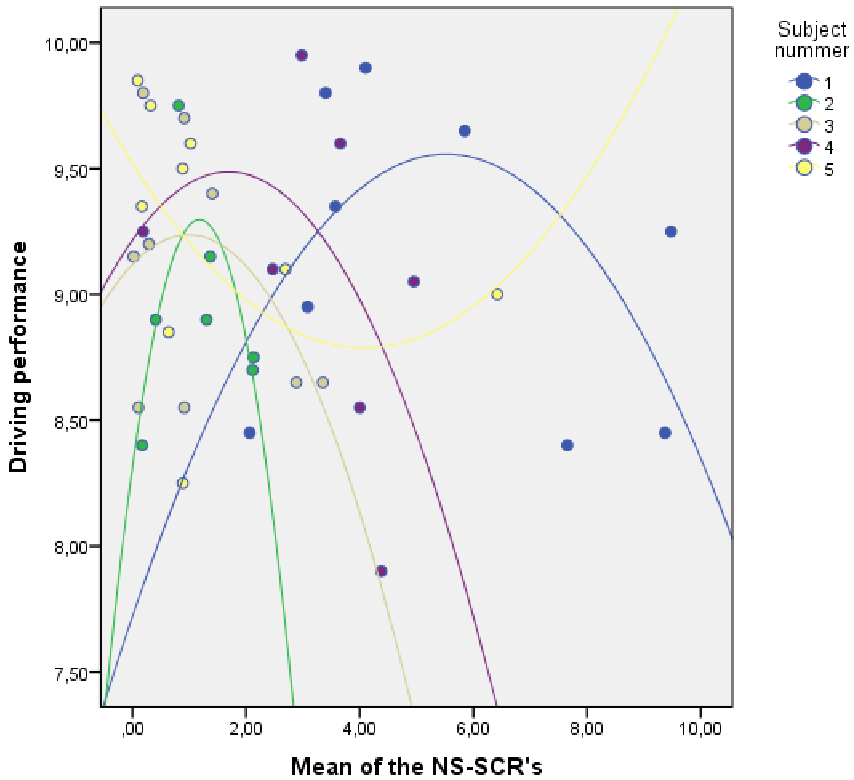


3.2 Performance

The reversed U shaped relation between performance and arousal suggests a quadratic regression between the performance and means of the NS-SCR variables. However the data does not support this. No significant effect was found for the mean of the NS-SCR's with the performance, t= -.28,*p* =.78. And the mean of the amplitude of the NS-SCR's, t= -.4,*p* =.69. The performance score for each participant can be found in appendix 1.

Looking at each individual there were no significant effect either between performance and the mean of NS-SCR, all *t*'s<2.1, all *p*'s> .11. The mean of the amplitude of NS-SCR's shows similar results, all *t*'s<1.1, all *p*'s> .35. When plotting the individual data, participants 1 to 4 show a reversed u shaped curve. As seen in figure 7.

Figure 7. Driving performance with mean of the NS-SCR for each participant (with quadratic regression line).



Another performance comparison is done with the NASA-TLX scores. The NASA-TLX scores can be found in appendix 2. Both NASA-TLX scores showed no significant relation with the performance. The mental demand with performance, t= 1.49 ,*p* =.15 and the frustration with performance, t= -.02,*p* =.99. At individual level there were nearly no significant effects. Looking at the mental demand with performance participant 1 had, t= -.63,*p* =.55, participant 2 had, t= -.69,*p* =.53, participant 3 had, t= -.17,*p* =.87, participant 4 had, t= 3.24,*p* =.03 and participant 5 had, t= .61,*p* =.56. And frustration with performance showed all *t'*s<1.4, all *p'*s> .22. Only one situation showed a significant effect. The relation between mental demand and performance was significant for participant 3.

3.3 NASA TLX

The NASA-TLX scores were compared with the means of the NS-SCR's variables. Mental demand with mean of total NS-SCR's (t= .03,*p* =.97) and with the mean amplitude of NS-SCR's(t =.3,*p* =.77). Also Frustration with mean of total NS-SCR's (t = -.96,*p* =.35) and with the mean amplitude of NS-SCR's(t =.23, *p* =.82). Both NASA-TLX scores showed no significant effect with the means of the NS-SCR variables.

Because the NASA-TLX is a subjective scale, there might be individual differences between the NS-SCR's variables and the NASA-TLX. So the same analysis was carried out for each individual. The results revealed no significant effect for nearly all the participants. With participant 3 there was a significant regression between the mental demand and the mean amplitude of NS-SCR's, t= 2.94,*p* =.026.

3.4 Simulator logging

The simulator data used in this analysis is the strip index, distance to next vehicle and distance to next element. This analysis will be based on the average per minute for each variable. Table 2 shows that the distance to preceding car does not correlate with any of the NS-SCR variables. There is however a significant effect between strip index, distance to next element and the two NS-SCR variables.

Table 2. Linear regression analysis of the NS-SCR's variables compared with strip index, distance to next vehicles and distance to next element.

	Total NS-SCR's		Amplitude of NS-SCR	
	t	Sig.	t	Sig.
Strip index	5.01	.01	1.66	.1
Distance to next vehicle	1.78	.08	.09	.93
Distance to next element	4.26	.01	5.85	.01

However, when looking at table 3 it is clear that the overall regression analysis does not speak for each individual. Participant 1 shows a significant effect for the total of the NS-SCR's with the strip index and distance to next element. But with the amplitude of the NS-SCR's only the distance to next element is significant. With participant 2 only the total of NS-SCR's with the strip index is significant. On the contrary participant shows a significant effect between all the variables. Participant 4 has a significant effect between the total of NS-SCR's and the strip index and with participant 5 the strip index has a significant effect on both NS-SCR variables. So the overall analysis does not speak for each participant.

Table 3. Linear regression analysis of the NS-SCR's variables compared with strip index, distance to next vehicles and distance to next element for each participant.

Participant		Total NS-SCR's		Amplitude of NS-SCR	
		t	Sig.	t	Sig.
1	Strip index	2.21	.03	-0.59	.56
	Distance to next vehicle	1.61	.11	-0.48	.63
	Distance to next element	3.81	.01	6.9	.01
2	Strip index	1.68	.09	1.28	.2
	Distance to next vehicle	.62	.53	-0.45	.65
	Distance to next element	2.34	.02	0.92	.36
3	Strip index	-2.65	.01	-2.07	.04
	Distance to next vehicle	-3.79	.01	-3.42	.01
	Distance to next element	5.15	.01	7.16	.01
4	Strip index	3.42	.01	1.2	.23
	Distance to next vehicle	1.14	.25	-0.08	.93
	Distance to next element	.16	.87	0.67	.5
5	Strip index	9.27	.01	4.91	.01
	Distance to next vehicle	1.33	.18	1.98	.05
	Distance to next element	0.73	.47	-0.30	.76

4. Discussion

The goal of this research was to observe novice drivers that follow the simulator driving course at a driving school. This course teaches the novice drivers the basics of driving a car and teaches them insight in modern day traffic. During the entire course the EDA levels were measured. This longitudinal aspect of the study made it possible to measure individuals over an extended period of time. During this period the participants were exposed to a virtual environment and they were learning new skills. The large sample of different measurements were investigated and showed some interesting results.

The first hypothesis is: Is there a relation between the EDA levels and the NASA TLX variables, Mental Demand and Frustration? Like arousal the NASA TLX test is a way to measure workload. They differ in that arousal is a physiological measurement and the NASA TLX is a subjective measurement (Rubio, et al., 2004). However when analyzing the NASA TLX and the EDA levels there were no effects found, except for participant 3. Participant 3 showed a significant effect between de mental demand subscale and the amplitude of the NS-SCR's. With the exception of participant 3, the subjective measurement doesn't have any connection with the physiologic measurement. In contradiction to Lee and Lui (2003) who found a correlation between the physiological measurement and the subjective measurement. While a big variation in the workload was measured through EDA, the participants did not express this in the NASA-TLX test. Making these results inconsistent with other studies (Y. H. Lee & Liu, 2003; Miyake, 2001).

The other three hypotheses are focused on the simulator log files. Three log files were compared with the EDA levels. Hypotheses two: Is there a relation between the EDA levels and the position of the participant on the road? The position on the road is determent with the strip index. When looking at the total data this is partially true. The total of NS-SCR's was significant, corresponding with previous studies that lane-variability does influence workload (Beede & Kass, 2006; Brookhuis, et al., 1991; Just, et al., 2008; Strayer & Johnston, 2001). However the amplitude of the NS-SCR's was not. At an individual level there are some differences. All participants except participant 2 showed a significant effect between the total of NS-SCR's and the strip index. As for the amplitude of the NS-SCR's participants 3 and 5 showed a significant effect. Although it was expected that there would be a significant relation between the EDA levels and the position on the road, this was only partly true. Participant's 3 and 5 showed a significant relation between both EDA measures. The other participants only had a significant relation with one of the EDA measures. These overall results are partly consistent with previous findings (Engström, et al., 2005; Jamson & Merat, 2005; Reimer, 2009).

The third hypothesis: Is there a relation between the EDA levels and the distance to the next vehicle? The distance to the next vehicle is expressed in meters, however this value is -1 when there is no vehicle in front of the driver. The overall data reveals no reason to suspect a relation between the EDA levels and the distance to the next vehicle. Showing a contradiction with the research by Schieβl, et al. (2006) who found an increase in cognitive demand when following a vehicle. Two individuals did show a significant effect. The EDA levels from participant 3 both related with the distance to the next vehicle, with participant 5 only the amplitude of the NS-SCR's.

The fourth hypothesis: Is there a relation between the EDA levels and the distance to the next element? These elements consist of road elements that a participant can encounter. Different types of roads are known to cause different levels of arousal (De Waard, et al., 1995). When looking at all the participants, the EDA levels revealed a significant effect for the distance to the next element. At an individual level there were some differences. Complex road types are more demanding than simple road types (Horberry, et al., 2006). Not all participants did experience a new road type as significantly more or less demanding. The total of NS-SCR's of Participants 1 and 2 was significant, the amplitude of the NS-SCR's was not. The EDA levels were both significant for Participant 3. Participants 4 and 5 showed no effects.

Looking at the individuals and the simulator log files there are some major differences. Participant 3 has significant effects on all the simulator variables, whereas participant 4 shows nearly no effect with the simulator variables. As stated in the introduction high demands on driving tasks cause an increase in skin conductance (Hoedemaeker & Brookhuis, 1998; Katzourakis, et al., 2011). The tasks that influence these variables can be experienced as cognitively demanding by some participants and therefore cause an increase in EDA levels. Whereas other participants do not experience it as cognitively demand. Participants differ in their experienced cognitive demand and therefore do not always show a significant relation between the EDA levels and the simulator variables.

Two different EDA measurements were taken during the research, giving insight in the EDA levels of the participants. These EDA levels are known to differ from person to person (Picard, 2009). This research was no exception, the total EDA levels of the participants differed from each other. Both the statistic test and the individual data revealed that there were some differences between the participants. Due to the longitudinal nature of this observation study it was possible to look for habituation effects. When looking at the differences between the EDA levels over time it clearly shows a decrease of the EDA level. Habituation is characterized by decreasing response intensity with repeated stimulation (Humphrey, 1933). This indicates that there is a habituation effect, although this habituation effect was not the same for each participant. Habituation can differ from person to person, Participants 1 and 3 showed the biggest habituation effect followed by participant 4 and participant 2 and 5 showed little habituation. This complies with earlier researches and illustrates once again EDA levels can differ from person to person (Howarth & Hodder, 2008).

The driving simulator scores participants on their driving performance. This is an overall indication of performance during each lesson. Yerkes and Dudson (1908) stated that there is a relation between arousal and performance. Although there was no linear effect found between the driving performance and EDA level. A similar result was found in another study (Vaez Mousavi, et al., 2011). The plot of the EDA level and diving performance does show a parabolic line that increases to an optimal point and then decreases. This is an effect that is described by the Yerkes and Dudson law. Performance increases with arousal when the arousal level is low, then reaches its peak at the optimal arousal level and then decreases as arousal continues to increase. This is shown with an inverse U-shaped curve. The Yerkes and Dudson law still applies here, as was found in several other studies (Ota, Toyoshima, & Yamauchi, 1996; Wrisberg, 1994).

4.1 General limitations

This study has some general limitations that need to be considered. This is due to the nature of this research, being an explorative ideographical study. The limited participant's size made it possible to do a longitudinal study in the amount of time that was available for a bachelor thesis. However this limits the generalization of the study, result of five participants are not representative enough for the population of novice drivers.

The EDA data consists of total NS-SCR's and the amplitude of NS-SCR's. By splitting up the EDA data it is difficult to interpret EDA as a whole. The data was also split up into one minute intervals. SCR can be related to a stimulus that occurred 5 seconds before the SCR. However with such large intervals it is not possible to make statements on specific situations or events.

Due to missing data files only the lesson 2 to 4 are missing from the dataset, this makes it difficult to look for habituation effect. Habituation can occur every time someone comes into contact with a stimulus, therefore the first lessons are important to analyze a habituation effect.

The EDA levels are only recorded during the driving lesson, this is a cognitive demanding task. Therefore there are no resting moments to compare the EDA levels with. No baselines where established for the participants.

4.2 Recommendations for further research

This explorative ideographical study creates possibilities for new follow-up studies. The goal was to explore behavior in a driving simulator through objective and subjective measures. This gave a general impression of EDA responses on a driving simulator environment. Some results were consistent with previous research, others were not. This led to new research options.

Individual differences are very much presented in the EDA data. When analyzing the EDA data and the other variables the results between participants were sometimes very different. Further analyzing of these differences can explain more on simulator behavior and show significant effects at an individual level. A way to make the EDA data more meaningful is by comparing the EDA levels in rest state with the EDA levels in the simulator. This can illustrate how much a participant reacts to a simulator environment in comparison with a normal rest state. This baseline increases the reliability of conclusion based on the EDA.

The NASA TLX and EDA are both measurements of workload, however they did not show any similarities with each other. This is an unexpected situation, because previous research has shown that there is a correlation between subjective and physiological measures of workload (Y. H. Lee & Liu, 2003). Further research may confirm or refute these findings. Also there are other subjective workload measurements that might show a different result with the EDA levels (Rubio, et al., 2004).

A comparison between reality and the virtual environment can give insights in the fidelity of driving simulators. Training in a experimental setting with a virtual environment and in reality show similar results (Rose, et al., 2000). However, comparing these in high stake situation can give new insights in the usability of a driving simulator.

4.3 Conclusion

The goal of this study was to observe novice drivers during their driving simulator course. Their behavior was observed with subjective and objective measures. Participants differ from each other in their EDA levels. The EDA levels habituated overtime, some participants showed a greater habituation effect then others. The NASA TLX test and EDA levels who measure the same thing had no similarities. EDA levels and performance showed an inverted u shape that is consistent with the Yerkes and Dudson Law. Performance and the NASA TLX had no significant effect apart from one participant. The simulator log files showed some effect with the EDA levels, this differed between participants. The results were not always consistent with what was expected. However this gives possibilities for further research.

Acknowledgments

This research project wasn't possible without the supervision and help of others. First I want to thank Matthijs Noordzij and Teun Lucassen for their guidance and supervision during this research project. I also want to thank “Auto en MotorRijschool Lucassen” for providing the space and de means to make the data collection possible. I also want to thank Jaap Lucassen for his cooperation in this research and support in operating the simulator. For the pleasant cooperation and support I want to thank Raphaela Schnittker. And I want to thank Green Dino for their open mindedness for this research

References

Allen, R. W., Park, G. D., Cook, M. L., & Fiorentino, D. (2007). The effect of driving simulator fidelity on training effectiveness. *Paper Presented at the Driving Simulation Conference North America*. doi: Cited By (since 1996) 1

Beede, K.E., & Kass, S.J. (2006). Engrossed in conversation: The impact of cell phones on simulated driving performance. *Accident Analysis & Prevention, 38*(2), 415-421.

Boucsein, W. (2011). *Electrodermal activity*: Springer Verlag.

Brookhuis, K.A., de Vries, G., & de Waard, D. (1991). The effects of mobile telephoning on driving performance. *Accident Analysis & Prevention, 23*(4), 309-316.

Cacioppo, J.T., Tassinary, L.G., & Berntson, G.G. (2007). *Handbook of psychophysiology*: Cambridge Univ Pr.

CentraalBureauRijvaardigheidsbewijzen. (2011). Centraal Bureau Rijvaardigheidsbewijzen Jaarverslag 2011, from http://www.cbr.nl/brochure/CBRjaarverslag2011.pdf

De Groot, S., De Winter, J. C. F., García, J. M. L., Mulder, M., & Wieringa, P. A. (2011). The effect of concurrent bandwidth feedback on learning the lane-keeping task in a driving simulator. *Human Factors, 53*(1), 50-62. doi: Cited By (since 1996) 6

De Waard, D., Jessurun, M., Steyvers, F. J. J. M., Raggatt, P. T. F., & Brookhuis, K. A. (1995). Effect of road layout and road environment on driving performance, drivers' physiology and road appreciation. *Ergonomics, 38*(7), 1395-1407.

Engström, J., Johansson, E., & Östlund, J. (2005). Effects of visual and cognitive load in real and simulated motorway driving. *Transportation Research Part F: Traffic Psychology and Behaviour, 8*(2 SPEC. ISS.), 97-120. doi: Cited By (since 1996) 79

GreenDino. (2012). Rijsimulator Retrieved 5-7, 2012, from http://www.greendino.nl/Classic.html

Hart, S.G., & Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Human mental workload, 1*, 139-183.

Hoedemaeker, M., & Brookhuis, K. A. (1998). Behavioural adaptation to driving with an adaptive cruise control (ACC). *Transportation Research Part F: Traffic Psychology and Behaviour, 1*(2), 95-106. doi: Cited By (since 1996) 45

Horberry, T., Anderson, J., Regan, M.A., Triggs, T.J., & Brown, J. (2006). Driver distraction: the effects of concurrent in-vehicle tasks, road environment complexity and age on driving performance. *Accident Analysis & Prevention, 38*(1), 185-191.

Howarth, P.A., & Hodder, S.G. (2008). Characteristics of habituation to motion in a virtual environment. *Displays, 29*(2), 117-123.

Humphrey, G. (1933). The nature of learning in its relation to the living system.

Jamson, A. H., & Merat, N. (2005). Surrogate in-vehicle information systems and driver behaviour: Effects of visual and cognitive load in simulated rural driving. *Transportation Research Part F: Traffic Psychology and Behaviour, 8*(2 SPEC. ISS.), 79-96. doi: Cited By (since 1996) 36

Jang, D. P., Kim, I. Y., Nam, S. W., Wiederhold, B. K., Wiederhold, M. D., & Kim, S. I. (2002). Analysis of physiological response to two virtual environments: Driving and flying simulation. *Cyberpsychology and Behavior, 5*(1), 11-18.

Johnson, L.C., & Lubin, A. (1966). Spontaneous electrodermal activity during waking and sleeping. *Psychophysiology, 3*(1), 8-17.

Just, M.A., Keller, T.A., & Cynkar, J. (2008). A decrease in brain activation associated with driving when listening to someone speak. *Brain research, 1205*, 70-80.

Katzourakis, D. I., Abbink, D. A., Happee, R., & Holweg, E. (2011). Steering force feedback for humanmachine-interface automotive experiments. *IEEE Transactions on Instrumentation and Measurement, 60*(1), 32-43. doi: Cited By (since 1996) 5

Kemeny, A., & Panerai, F. (2003). Evaluating perception in driving simulation experiments. *Trends in Cognitive Sciences, 7*(1), 31-37. doi: Cited By (since 1996) 77

Lee, W.S., Kim, J.H., & Cho, J.H. (1998). *A driving simulator as a virtual reality tool*.

Lee, Y.H., & Liu, B.S. (2003). Inflight workload assessment: Comparison of subjective and physiological measurements. *Aviation, space, and environmental medicine, 74*(10), 1078-1084.

Malmo, R.B. (1959). Activation: A neuropsychological dimension. *Psychological Review, 66*(6), 367.

Mantovani, F. (2001). VR Learning: Potential and Challenges for the Use of 3D Environments in Education and Training. *Towards cyberpsychology: mind, cognition, and society in the Internet age, 2*(1ntroduction), 207.

Mehler, B., Reimer, B., Coughlin, J.F., & Dusek, J.A. (2009). Impact of incremental increases in cognitive workload on physiological arousal and performance in young adult drivers. *Transportation Research Record: Journal of the Transportation Research Board, 2138*(-1), 6-12.

Meshkati, N., Hancock, P.A., Rahimi, M., & Dawes, S.M. (1995). Techniques in mental workload assessment.

Miyake, S. (2001). Multivariate workload evaluation combining physiological and subjective measures. *International journal of psychophysiology, 40*(3), 233-238.

Nagai, Y., Critchley, HD, Featherstone, E., Trimble, MR, & Dolan, RJ. (2004). Activity in ventromedial prefrontal cortex covaries with sympathetic skin conductance level: a physiological account of a “default mode” of brain function. *Neuroimage, 22*(1), 243-251.

Ota, T., Toyoshima, R., & Yamauchi, T. (1996). Measurements by biphasic changes of the alpha band amplitude as indicators of arousal level. *International journal of psychophysiology, 24*(1), 25-37.

Picard, R.W. (2009). Future affective technology for autism and emotion communication. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*(1535), 3575-3584.

Poh, M.Z., Swenson, N.C., & Picard, R.W. (2010). A wearable sensor for unobtrusive, long-term assessment of electrodermal activity. *Biomedical Engineering, IEEE Transactions on, 57*(5), 1243-1252.

Psotka, J. (1996). Immersive training systems: Virtual reality and education and training. *Instructional Science, 23*(5-6), 405-431.

Reimer, B. (2009). Impact of cognitive task complexity on drivers' visual tunneling. *Transportation Research Record*(2138), 13-19. doi: Cited By (since 1996) 16

Reimer, B., Mehler, B., Coughlin, J.F., Godfrey, K.M., & Tan, C. (2009). *An on-road assessment of the impact of cognitive workload on physiological arousal in young adult drivers*.

Rose, FD, Attree, EA, Brooks, BM, Parslow, DM, & Penn, PR. (2000). Training in virtual environments: transfer to real world tasks and equivalence to real task training. *Ergonomics, 43*(4), 494-511.

Rubio, S., Díaz, E., Martín, J., & Puente, J.M. (2004). Evaluation of Subjective Mental Workload: A Comparison of SWAT, NASA-TLX, and Workload Profile Methods. *Applied Psychology, 53*(1), 61-86.

Schießl, C., Vollrath, M., Altmüller, T., Dambier, M., & Kornblum, C. (2006). *What is strenuous? Driving itself or the driving situation*.

Steiner, G. Z., & Barry, R. J. (2011). Exploring the mechanism of dishabituation. *Neurobiology of Learning and Memory, 95*(4), 461-466.

Strayer, D.L., & Johnston, W.A. (2001). Driven to distraction: Dual-task studies of simulated driving and conversing on a cellular telephone. *Psychological science*, *12*(6), 462-466.

Vaez Mousavi, S. M. K., Naji, M., Osanlou, M., & Esmaeilpour Marandi, H. (2011). Cadets' performance prediction in a static balance task based on arousal and activation. *Journal of Military Medicine*, *13*(2), 109-116.

Wrisberg, C.A. (1994). The arousal-performance relationship. *Quest (National Association for Physical Education in Higher Education), 46*(1), 60-77.

Yerkes, R.M., & Dodson, J.D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of comparative neurology and psychology*, *18*(5), 459-482.

Appendix

Appendix 1 – NASA-TLX scores on the variables mental demand and frustration for each participant.

			Participant				
	Lesson	NASA_TLX scale	1	2	3	4	5
sessie	1	Mental Demand	7.4	6.5	11.0	3.1	11.0
		Frustration	4.3	.0	6.4	19.3	7.6
	2	Mental Demand	6.9	-	9.3	10.2	12.2
		Frustration	4.3	-	1.2	.0	10.4
	3	Mental Demand	5.2	-	17.1	8.3	17.8
		Frustration	1.1	-	4.7	7.6	20.8
	4	Mental Demand	6.2	9.8	1.3	-	21.3
		Frustration	3.1	0.0	3.3	-	29.7
	5	Mental Demand	5.9	5.2	5.3	16.3	14.4
		Frustration	1.6	.0	3.8	0.0	28.3
	6	Mental Demand	3.2	6.5	2.7	14.4	25.9
		Frustration	1.1	0.0	3.0	0.0	7.1
	7	Mental Demand	5.2	9.8	1.8	-	32.0
		Frustration	1.7	1.7	2.4	-	18.4
	8	Mental Demand	2.3	10.0	2.0	19.0	30.7
		Frustration	.9	.0	1.8	.0	11.0
	9	Mental Demand	4.5	3.3	2.0	26.7	26.0
		Frustration	2.9	0.3	17.3	5.2	26.7

Appendix 2 – Driving performance scores for each participant separated per lesson.

	Participant				
	1	2	3	4	5
Lesson					
1	8.45	8.7	8.65	7.9	9.1
2	8.95	-	8.65	8.55	8.25
3	9.25	-	9.4	9.05	9.5
4	9.35	8.9	9.8	-	8.85
5	9.65	8.75	9.2	9.6	9.85
6	8.45	8.4	8.55	9.1	9.75
7	8.4	8.9	8.55	-	9.00
8	9.9	9.15	9.15	9.25	9.35
9	9.8	9.75	9.7	9.95	9.6