An idiographic study into the physiology and self-reported mental workload of learning to drive a car

Matthijs L. Noordzij1, Serena M. Dorrestijn1, & Irma A. van den Berg2
1University of Twente
2Irma van den Berg Tekst- en verkeersproducties
The Netherlands

Abstract

A driving instructor has to monitor the performance and state (e.g. mental workload) of the pupil who is learning to drive. However, the instructor is also responsible for road safety. Therefore, it might be beneficial when additional monitoring technology would be available to the driving instructor. Fluctuations in skin conductance are indicative of changes in the autonomic nervous system and have been operationalized as changes in stress or mental workload. For the present idiographic study six pupils were followed during their driving training, while measuring their self-reported (and by their driving instructor) workload and their skin conductance levels (with a wrist-worn bio-sensor). The quality of the physiological measurements was acceptable in most cases. Most students showed their highest physiological values 3-7 lessons before their final exam. The driving instructor was good at predicting the self-reported workload of her pupils. Importantly, there was no correlation between physiological fluctuations and fluctuations in self-reported workload. This makes skin conductance measurement unsuitable to replace subjective workload assessments. The physiological data did provide first evidence that a modular driving instruction methodology, with several partial exams, seems to prevent extremely high physiological activity during the final exam.

Introduction

A driving instructor has to closely monitor elements in the environment and in the car related to road safety, as well as the learning progress and capacity of the pupil. These tasks are performed up to 8 hours a day and can be considered demanding monitoring tasks where the instructor has to stay vigilant. It is well known that performance on vigilance and monitoring tasks can strongly diminish after relatively short times (under 30 minutes) (Warm, Parasuraman, & Matthews, 2008). At present, wearable (i.e. applicable on the body) technology methods are becoming available to continuously monitor physiological measures (e.g. heart rate), which might be helpful for a driving instructor to keep insight into the physical or mental state of the pupil. However, it is unclear whether these measures can, to some extent, replace the personal assessment of an instructor or if these measures are actually complementary and might provide previously unavailable, continuous insight into a student’s state. The present study examines the patterns of self-reported workload.
and physiological measurement from a new type of wrist-worn bio sensor which measures electrodermal activity (EDA) during real life driving lessons.

A construct that is often used in Human Factors is mental workload. As a first approximation mental workload can be described as academic jargon for ‘mental effort’. It is situated within the information processing model of human cognition (Vidulich et al., 2010) and is defined by Parasuraman et al. (2008) as “the function relating the mental resources demanded by a task and those resources available to be supplied by the human operator” (pp. 145–146). For decades researchers have been studying the mental workload of drivers. It is beyond the scope of this paper to exhaustively introduce theories on mental workload and how fluctuations of mental workload occur in different driving situations (for excellent reviews see for example De Waard, 1996, and Vidulich et al., 2010). It is clear that learning to drive is a complex task with many different demands, which are placed on the novice driver at the same time. During driving unexpected events that have to be dealt with immediately occur often (e.g. traffic jams, road blocks and detours). Taken together it is clear that learning to drive relies heavily on perceptual and cognitive abilities, and that risks of cognitive overload are greater for student and novice drivers (Ross et al., 2012). Therefore, measures of workload during the crucial phase of learning to drive a car could be very welcome to both driving instructor and student.

A simple, but reliable and valid (Verwey & Veltman, 1996) way to measure self-reported workload, is via the Rating Scale of Mental Effort (RSME) (Zijlstra, 1993). This unidimensional scale asks participants to rate the effort they expanded on a task they have just carried out. Our study is the first, to our knowledge, that follows student drivers in a longitudinal manner through several driving lessons and their exam, and measures RSME values for lessons. This makes it difficult to make precise predictions on how student drivers and their instructors will report on the mental effort of the lessons. Our first exploratory research question pertains to the type of patterns we will find for self-reported mental workload during driving lessons. Although precise predictions are difficult we would like to make two suggestions based on (1) literature on novice drivers and (2) the type of driving training our participants receive. First, it is well known that over time people automate the visuo-motor control aspects of driving (e.g. shifting gear), as well as habituate to certain driving situations and scenario’s (e.g. Verwey, 2000). From this it could be expected that students will show declining levels of self reported mental workload over lessons. Second, the present study follows student drivers who are being instructed according to a modular driving instruction methodology (‘Rijopleiding in Stappen’). A traditional, non-modular system (still most common in the Netherlands) would have a final exam in which effort and stress levels are expected to be very high. In contrast, the modular system tries to shift this moment with several partial exams. When approximately 75% of the lessons are finished most of the required skills can be shown and passed in a final partial exam. The actual final exam is presented more as a formality, which all students are expected to pass. From this it could be predicted that students might show their highest mental effort 3-7 lessons prior to the final exam. Next to self-reported measures this study also includes EDA as a possible physiological index of workload.
EDA refers to all electrical phenomena in the skin, most often expressed in skin conductance (SC) units (Boucsein, 2012), which are typically measured at the fingers or the hand palm. The SC signal can be described as small, short waves (Skin Conductance Responses (SCRs)) riding on a larger wave (the Skin Conductance Level (SCL)). Fluctuations in the occurrence and amplitude of SCRs are caused by activity of the eccrine sweat glands, which are innervated purely by the sympathetic nervous system. Typical values (mostly from lab studies) are between 0 to 20 SCRs per minute and an average SCL of 2 µSiemens (µS) (Dawson, Schell, & Filion, 2007).

EDA has been systematically applied in traffic research (Boucsein, 2012). It is often mentioned as one of the methods to do online physiological approximations of workload, but consensus is that it is a rather general measure of arousal or stress (e.g. Healey & Picard, 2005). Recently, it has become possible to measure EDA at the wrist in an ambulatory non-intrusive fashion with no wires, thereby minimizing motion artifacts (Poh, Swenson, & Picard, 2010). The second research question is what the pattern of skin conductance parameters over lessons will be from a wrist-worn EDA sensor. Similar to the RSME values we will explore the data over time for the presence of consistent patterns.

Next to the central research questions concerning the possible RSME and skin conductance patterns during driving lessons three related side issues will also be explored. (1) The standard location to measure EDA is the fingertips or the hand palm. For this study we will measure EDA at the wrist. Although this study is by no means a validation study for the wrist-worn sensor we will be able to find out whether relevant skin conductance parameters (i.e. number and amplitude of SCRs and the SCL) can be extracted from data from a wrist-worn bio sensor during driving lessons. (2) In addition we can test whether any of the skin conductance parameters correlate with self-reported RSME values. (3) Finally the driving instructor will also assess how mentally effortful a lesson was for her student. A previous study had suggested that self-report and report by a driving instructor had medium correlations (Victoir et al., 2005). We will try to replicate this finding in our study.

To explore the research questions an idiographic approach (i.e. the science of individualized measurements) is employed (Molenaar, 2004). Instead of measuring data for a great number of student drivers for one lesson, a limited number of students is followed during a number of lessons (including their exam). There are several reasons why such an idiographic approach is appropriate for our research questions. The individual differences in relation to fluctuations in both self-reported measures and physiology are large, but individuals might have clear patterns over time. Moreover, our research is exploratory and would benefit from extensive individual data, which subsequently could inform more hypotheses driven nomothetic studies (e.g. studies in which group means are compared such as a randomized controlled trial). Finally, as stated by Picard (2009, p. 3580), and potentially very relevant for individuals looking for insight in the physical and mental effort during driving lessons: ‘with an individualized data-intensive approach based on measurement in a person’s natural environment it is not just the science
that benefits: each participant can now benefit with information specific to his or her
needs and situation.’

**Method**

**Participants**

Six students (four female and two male aged between 17 and 22 years) from driving
school Irma van den Berg (www.irmavandenberg.nl) were recruited via convenience
sampling.

**Materials**

**Q sensor**

The Affectiva Q™ sensor is a wrist worn, watch-like sensor which measures EDA
with 1cm diameter Ag-AgCl dry electrodes at the ventral side of the wrist. In
addition to EDA, actigraphy and skin surface temperature data are also logged at a
sampling rate of 32Hz.

**Rating Scale of Mental Effort (RSME)**

The RSME (Zijlstra, 1993) is a unidimensional, reliable instrument to measure self-
reported mental workload. The RSME is a 15 cm long line (150-points) with
markings at every centimetre, and nine anchor points with verbal labels going from
“absolutely no effort” (around the 0 point) to “extreme effort” (around the 112th
point on the scale). Answers can be given by marking the line at a point that fits the
experienced workload.

**Design & Procedure**

All students (or their legal caretakers for students younger than 18 years) signed an
informed consent. All participants were explicitly told that the measurements would
have no influence on their driving lessons and that they were free to remove the bio
sensor at any time. In addition, a female driving instructor (57 Years old, 27 years’
experience as a driving instructor) acted as instructor during all the lessons.

As mentioned above, the present field study has an idiographic and observational
approach and design. A limited number of driving students were followed
intensively during (parts of their) driving lessons (including their driving exam) in
the same Mitsubishi Outlander with a manual gear box. The driving lessons occurred
as they would have normally, except for the following differences: 1) at the start and
at the end of the lesson both the student and the instructor filled in the RSME. The
student was asked to fill in the RSME at the start to indicate how effortful the task
was he or she had been doing before the driving lesson began (not used for this
present paper), and to indicate on the RSME at the end of the how effortful the
driving lesson as a whole had been. The instructor was asked to fill in the RSME
estimating the effort of the student during the driving lesson (thus not her own
effort). The student and instructor were blind to each other’s rating. 2) At the start of
the lesson the Q sensor was attached to the left wrist by the student with the
instruction to get a tight, but comfortable fit (using an easy Velcro band).
The left wrist was chosen to minimize the influence of movement, which was deemed more likely on the right wrist due to movements related to (manual) gear changes. The left hand is mainly involved in smooth and slow turning motions probably making the influence of motions on the physiological measurements rather negligible (Kappeler-Setz et al., 2013). An exception might be special manoeuvres which require a quicker steering. Below (under the heading ‘Motion Artefacts’) a description is provided on how the actigraphic data was correlated to the EDA data to check whether motion fluctuations are associated with EDA parameter fluctuations.

No pressure was put on the instructor to use the Q sensor during any given lesson or the exam. Furthermore, there was no experimenter present in the car during the lesson or the exam, making the driving situation identical to a normal driving lesson or exam.

Data analysis

Skin Conductance parameters

EDA data were down sampled to 16Hz, and pre-processed with a Continuous Decomposition Analysis (CDA) as implemented in Ledalab (Benedek & Kaernbach, 2010), which requires MATLAB (www.mathworks.com). From the CDA an estimate of the skin conductance level (SCL) was acquired. The phasic activity coming from classical Through-to-Peak analysis was reported (threshold for an SCR amplitude was set at .03 µS) (Boucsein, 2012).

As recommended (Boucsein, 2012), visual checks were performed on plots of skin conductance data to identify failed measurements, “non-responding” (indicated by an absence of SCRs in a given measurement), and incorrect classification of SCRs. Data from these problematic measurements were removed from further analysis.

The SCL and SCR parameters were expressed in a number of variables both at the minute level and at the lesson level. For every individual minute the mean SCL, number of SCRs and total amplitude of SCRs were calculated. In addition, mean SCL, mean number of SCRs per minute, and mean total amplitude per minute were calculated for every lesson. By also aggregating the skin conductance parameters per lesson, the physiological and self-reported RSME scores could be correlated, because the latter only concerns a complete lesson. Only the RSME score provided at the end of the lesson was taken into consideration.

Motion artefacts

To check whether fluctuations in motion were associated with fluctuations in skin conductance parameters, the total magnitude of the vector of acceleration (total acceleration per minute $t_a$) was calculated from the actigraphic data via $t_a = \sqrt{a_x^2 + a_y^2 + a_z^2}$. Subsequently, $t_a$ was correlated with the skin conductance variables (per minute).
Results

Quality of the EDA data and motion artefacts

We first checked whether the wrist worn sensor delivered usable data and whether the EDA parameters were associated with motion fluctuations. The Q sensor was applied during 99 lessons of which 86 (85%) resulted in EDA measurements that were suitable for subsequent analyses (Figure 1A). Total EDA measurement time was 7988 minutes, and the mean lesson time was 86 (SD. = 17.5) minutes with a range between 30 and 147 minutes. In Figure 1 EDA measurements are shown for two driving lessons (panel A and Panel B visualize the EDA data for two complete lessons). Most lessons were associated with typical EDA patterns (Figure 1A). Some lessons had failed measurements with no clear changes in skin conductance (Figure 1B, 15% of the data).

![Figure 1. A) An example of a typical successful measurement during a driving lesson depicting both variation in SCL and SCRs. B) A failed measurement due to “non-responding” of the participant. On the x-axis is Time (minutes), and on the y-axis is skin conductance level (μSiemens).](image)

There were no correlations between fluctuations in the total acceleration \(t_a\) and skin conductance variables calculated per minute (see Table 1). In contrast, as can be expected, the skin conductance variables showed strong to very strong correlations to one another.

<table>
<thead>
<tr>
<th></th>
<th>Acceleration</th>
<th>Number of SCR</th>
<th>Amplitude (SCR)</th>
<th>SCL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acceleration</td>
<td>-</td>
<td>-0.02</td>
<td>0.004</td>
<td>0.07*</td>
</tr>
<tr>
<td>Number SCR</td>
<td>-0.02</td>
<td>-</td>
<td>0.86*</td>
<td>0.67*</td>
</tr>
<tr>
<td>Amplitude SCR</td>
<td>0.004</td>
<td>0.86*</td>
<td>-</td>
<td>0.57*</td>
</tr>
<tr>
<td>SCL</td>
<td>0.07*</td>
<td>0.67*</td>
<td>0.57*</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 1. Correlations for acceleration and skin conductance variables per minute (all p’s < 0.001 marked with *, n (minutes) = 7277).
**RSME values and Skin conductance parameters: Possible patterns over lessons**

Figure 2 gives insight in how skin conductance parameters (Fig. 2 A and 2B) and RSME values (Figure 2C) fluctuate over lessons and between individuals. Our two research questions concerned patterns in RSME or SC values. There were strong variations in RSME values between lessons and students (see Figure 2C). Overall there was a positive correlation between lesson number and RSME values, $r(93) = .37$, $p < .001$, yet, this correlation differed strongly between students, from $r = -.68$ to .42. For almost all participants (one participant had a correlation of $r = -.1$) there were weak, to medium, positive correlations between lesson number and the skin conductance parameters, all $r$’s between .05 and .48, all $p$’s < .001.

![Figure 2. Variations of mean number of SCRs(A), mean SCL (µSiemens) (B), and mean RSME values per lesson (C), per student. The data for the mean amplitude of SCRs are not depicted because it follows a very similar pattern to the number of SCRs data (see also the high correlations in Table 1).](image_url)
Participants 1, 2, 4 and 5 showed their highest skin conductance values in the 3-7 lessons prior to their driving exam. Importantly, five out of six students passed their driving exam (student 6 stopped taking driving lessons altogether). Student 3 was no longer followed with the Q sensor after lesson 18 because of very low SCR values (virtually non-responding), and only the final exam was measured (so for 4 participants we had a complete data set).

We also tried to replicate earlier research that found medium correlations between self-reported workload by the student driver and the driving instructor. The evaluations by the instructor of the experienced workload of her students in our study correlated positively with the evaluations of the students themselves, $r(70) = .8, p < .001$ (range = .14 to .89, see Table 2).

Table 2. Descriptive statistics for RSME values as reported by students themselves and as assessed by the instructor. The number of lessons (n) with valid RSME values is indicated after the student number. The RSME exam column indicates what value students gave for their driving exam (student 1 did not rate the exam and student 6 stopped taking driving lessons).

<table>
<thead>
<tr>
<th>Student number</th>
<th>Range student</th>
<th>Mean student (sd.)</th>
<th>RSME exam</th>
<th>Range instructor</th>
<th>Mean instructor (sd.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (n = 20)</td>
<td>13 - 102</td>
<td>52 (30.1)</td>
<td>-</td>
<td>13 - 108</td>
<td>59 (36.4)</td>
</tr>
<tr>
<td>2 (n = 7)</td>
<td>40 - 110</td>
<td>83 (22.9)</td>
<td>70</td>
<td>27 - 118</td>
<td>86 (32.5)</td>
</tr>
<tr>
<td>3 (n = 9)</td>
<td>27 - 90</td>
<td>49 (17.3)</td>
<td>90</td>
<td>50 - 101</td>
<td>66 (21.5)</td>
</tr>
<tr>
<td>4 (n = 21)</td>
<td>13 - 86</td>
<td>41 (20.7)</td>
<td>72</td>
<td>13 - 101</td>
<td>43 (25.6)</td>
</tr>
<tr>
<td>5 (n = 12)</td>
<td>13 - 57</td>
<td>34 (11.5)</td>
<td>13</td>
<td>10 - 71</td>
<td>40 (22.3)</td>
</tr>
<tr>
<td>6 (n = 16)</td>
<td>13 - 38</td>
<td>20 (8.2)</td>
<td>-</td>
<td>10 - 57</td>
<td>37 (16.1)</td>
</tr>
</tbody>
</table>

Finally, the possible correlation between the reported RSME values and the physiological patterns were explored. Overall the fluctuations per lesson in self reported RSME values by the student and skin conductance parameters did not correlate, all $r's(85)$ between .01 and .05, all $p's > .66$. In addition, the RSME values indicated by the instructor also showed no correlation with the skin conductance parameters of the student, all $r's(65)$ between -.13 and -.03, all $p's > .3$. This total lack of correlations was also found when only the last ten minutes of SC data (the minutes close to the RSME measurement moment) of every lesson were taken into account, all $r's(85)$ between -.09 and -.05, all $p's > .43$. For one individual student (number 4) a correlation between mean number of SCR per minute and RSME values was significant. However, this correlation was negative, $r(22) = - .49, p = .03$.

Discussion

This idiographic study set out to explore physiological and self-reported indices of workload of young people learning to drive. The main results were that (1) relevant skin conductance parameters (with biologically plausible values) could often be extracted successfully from EDA signals acquired from a wrist-worn sensor during a real life driving lesson or exam. (2) Correlations showed that driving students’ self-
reported workload and skin conductance parameters increased, instead of decreased as the driving lessons progressed. However, for the self-reported workload this was highly student dependent with some students showing a negative correlation. In addition, visual inspection seemed to suggest that students reached their (mean) peak physiological values 3-7 lessons before the driving exam after which their values started to decrease. (3) There was absolutely no correlation whatsoever between self-reported (both by the student and the instructor) workload and skin conductance parameters. Below further elaboration will be provided on these findings in the light of the feasibility of ambulatory EDA measurements, Human Factors theory, and practical recommendations for the integration of EDA measurements during driving lessons.

EDA measurements at the wrist with dry electrodes seem technically possible and easy in a real-world driving lesson scenario. At present, the state of the art concerning EDA measurements comprises of measurements at fingers or in the palm with electrode gel (Boucsein et al., 2012). Our data showed that in 85% of the cases the EDA data patterns were according to what can be expected with clearly identifiable SCRs and very little motion artefacts. There were individuals that had quite a few non-responding sessions, but it must be noted that even when following the state of the art there will always be participants who exhibit a total lack of SCRs (Boucsein, 2012). The range of the number of SCRs per minute, and the average SCL were in a biologically plausible range (between 0-20 per minute, and near 2 µS, Dawson et al. (2007)). In addition, fluctuations in movement as captured by the accelerometer did not correlate at all with fluctuations in skin conductance parameters, providing evidence that motion artefacts were not a major problem in this case. The EDA signals in this study were not directly compared to a measurement at the fingers. However, other studies have done this comparison in a lab setting and showed that the wrist measurements are very comparable to the finger measurements (Poh et al., 2010). The present study adds to this lab validation a field validation of EDA wrist measurement in a particular context: Learning (how) to drive in a naturalistic setting.

Skin conductance parameters and RSME values (Zijlstra, 1993) did not steadily decline over the course of various driving lessons. The physiological values actually increased until shortly before the driving exam, after which they seemed to decrease. Decades of research into (cognitive) workload and habituation during driving lead to the idea that the repetition of an activity could lead to initial high values associated with a very explicit knowledge driven phase, after which more automatic cognitive-visuo-motor associations would form leading to less workload (e.g. Collet et al., 2009; De Waard, 1996; Dijksterhuis et al., 2011; Verwey, 2000). The present study suggests that the knowledge obtained from (novice) drivers might not be completely generalizable to student drivers. Up to 3-7 lessons prior to their (successful) driving exams skin conductance parameters increased. Only then did they start to decrease. This pattern in the skin conductance parameters fits with the basic assumption from the modular driving methodology the students were following, where the most important (partial) test is taken when 75% of the lessons have been completed. The goal is to make the final exam less stressful and more of a formality. The physiological data provide evidence that this is indeed working. The RSME values
were less clear with some students still showing the highest value for their final exam. Obviously, this value of the modular methodology needs to be verified with a larger, and importantly a more diverse sample of students passing, but also failing their exam, while following different driving instruction methodologies. However, it is exactly this kind of interesting pattern one can identify within an idiographic study, and subsequently test in larger (and more expensive) idiographic and also (perhaps at a later time) nomothetic studies.

It was found that the instructor was very good at predicting the RSME values of the participant. This is in accordance with previously reported correlations (Victoir et al., 2005) between instructor and student, although the strength of the correlation was much higher in our study (.8 vs. .4). Given that this particular driving school has a very high success rate, it would be interesting to examine the association between the instructor and the student in relation to student success in future studies.

There was no relation between fluctuations in self-reported workload and skin conductance parameters. Therefore, the present results fit within a theoretical position of emotion research which holds that self-reported measures are associated with and predictive of cognitive schemata, personality dimensions and recall biases, and not with physiological activation (Myrtek et al., 2005; Robinson & Clore, 2002). One could argue that the lack of a correlation is caused by the fact that we averaged the physiology over the whole lesson to compare it to the RSME value. However, only taking the average of the last minutes of the lesson and compare these to the RSME values did nothing to improve the correlation. In Human Factors literature this dissociation between self-reported workload and physiology has also been reported. It has been suggested that these two types of measures are indicative of different mechanisms and processes underlying fluctuations in mental workload (Johannes & Gaillard, 2014; Yeh & Wickens, 1988). Finally, as mentioned in the introduction, EDA as a physiological measure is not assumed to be selective for workload or mental effort (Boucsein, 2012). Therefore, the lack of a correlation for all skin conductance parameters to the RSME value, corroborate in a longitudinal ambulatory context, that EDA is unsuitable to replace specific subjective workload assessments (see also Seitz et al. (2013) for a similar result in experienced truck drivers).

Future research could focus more on the particular strengths of physiological measurements. Unlike self-report, physiological measurements are online, and continuous, providing insight into moment-by-moment fluctuations in arousal. It would be interesting to find out whether EDA can be a valid indicator of “peak” moments (De Waard, 1996), when arousal has become too high for optimal, or at least, acceptable driving and learning behavior. As mentioned above, different driving instruction methodologies should be systematically compared in terms of the physiological arousal that is associated with the various exam moments. It could be that more modular approaches (with several partial exams) are better at removing extreme stress from a final exam moment, and provide a more valid test of the actual driving skills.

This study provides some basis for recommendations to driving instructors. First, the RSME values could be collected quite easily after each lesson, and can be taken as a
rough estimate on what kind of learning trajectory the student are facing according to their own experience. In this study the driving instructor was overall very good in assessing the subjective state of her student (a correlation of .8 was found). However, this was not true for every student, and might be caused by specific qualities in this particular instructor. All instructors therefore could consider asking for the RSME value after each lesson to get the self-reported mental effort directly from the student. This would provide the instructor with direct insight into how much effort the student is putting into the lesson. Second, ambulatory EDA measurements are not easily available to the professional market. However, it seems this is about to change (see for example the “Embrace” hardware from Empatica, www.empatica.com/product-embrace). Instructors should be careful in using this hardware and interpreting the results as specifically indicative of mental effort. In general, EDA measurements should not be interpreted in a very specific way, but more as an additional information source (at present not easily available) about the arousal of students. At present, self-reported RSME ratings would give a better idea about the mental workload of student drivers.

Conclusion

The present study showed that closely following the physiology and self-reported workload of student drivers over time provides new and fresh insights. More research is needed on the questions whether physiological measurements would be valuable to use during driver training, and how wearable technology might play a role in this process.

References


