Advancing Simulation-Based Driver Training: Lessons Learned and Future Perspectives

J.C.F. de Winter, S. de Groot, J. Dankelman, P.A. Wieringa
BioMedical Engineering Department
Mechanical, Maritime and Materials Engineering
Delft University of Technology
Mekelweg 2, 2628 CD Delft, The Netherlands
+31 15 278 2156
j.c.f.dewinter@tudelft.nl

M.M. van Paassen, M. Mulder
Control and Simulation Department
Aerospace Engineering
Delft University of Technology
P.O. Box 5058, 2600 GB Delft
The Netherlands

ABSTRACT
This paper aims to provide recommendations for improving the effectiveness of automatic, student-adaptive, simulation-based driver training. Using experiments and recorded data in driving simulators, three distinct issues are discussed: 1) the student, 2) the virtual driving instructor (VDI), and 3) the student-profile. We found that: first, students seek task-relevant information themselves; not providing them with feedback can be beneficial. Second, an intelligent VDI that emulates a human driving instructor is not favored. To the contrary, regressive instruction – a relatively simple principle – was effective in letting students drive away autonomously. Third, constructing a student-profile based on individual characteristics, such as a strength-weakness report, is viable for providing student-adaptive feedback.

Categories and Subject Descriptors

General Terms
Measurement, Performance, Experimentation, Human Factors.

1. INTRODUCTION
The greater part of current driver training takes place on the road under the supervision of a human instructor. This traditional form of training is expensive, while research has shown that it does not reduce post-license crash risk as compared to informal training [1]. Driving simulators are a complementary tool to on-road training and offer advantages such as objective student assessment, standardization, free control over the training conditions, potential cost-effectiveness due to automation, and didactic possibilities such as multimodal feedback, demonstrations, and replays [2][3]. However, research and development of simulators is primarily hardware-oriented (e.g., developing display techniques), whereas the aforementioned advantages are not optimally exploited [2]. It is argued that research is needed to find out how to get the most out of simulators so that students are efficiently trained to obtain their license and to drive safely.

At present, the Netherlands has an important role in the domain of simulation-based driver training [4]. A major player is Green Dino Virtual Realities. Their Dutch Driving Simulator (DDS, [5]) is predominantly used for automatic training at driving schools across the country. In this training mode, a virtual driving instructor (VDI) provides feedback and instructions during the training sessions [6]. A human supervisor has the possibility to evaluate a student’s progress after a simulator-session has been completed and accordingly, to decide to alter the training curriculum, to repeat a lesson, or to transfer to a real car.

This paper aims to provide recommendations of how to improve the effectiveness of automatic, student-adaptive, simulation-based driver training systems. The focus of this paper is not on technical hardware requirements, but on didactic software requirements instead. Using experiments and recorded data in driving simulators, three distinct issues are discussed:

The student. Considering that humans are intelligent systems, it is first explored what students can do by themselves and when they actually need feedback and instructions.

The virtual driving instructor (VDI). It is investigated how the VDI can be improved. Should the VDI’s intelligence be increased so that it better understands the human student? The VDI’s effectiveness is studied in a specific task (driving away) and suggestions of improvement are provided.

Constructing a student-profile. Herein, we investigate whether a student-profile based on individual differences in driver behavior can be used for advancing training effectiveness.

Even though a simulator is a stationary system (providing an illusion of being on the move), the results in the present study bear direct relevance to human-computer interaction with mobile devices. A student-profile is portable and can possibly be used in advanced driver assistance systems (ADAS) or future automotive navigation systems.

2. THE STUDENT

2.1 Self-paced task
Humans are active information seekers rather than passive recipients. This presumption also applies to simulation-based driver training, since car driving is, to a large extent, self-paced.
Previous research showed that there were considerable differences between students in driving simulation performance, with respect not only to success rates on driving tasks, but also to driving speed and speed of task execution [8]. To illustrate, Figure 1 shows students’ (N=1760) time to complete a lap around a square block of intersections during a ‘turn right’ exercise using computerized driving instruction. There was no other traffic. The data have been obtained by DDSs stationed at driving schools in the Netherlands in the period August 2005 – March 2006. It can be seen that there were considerable individual differences in driving speed, indicating that students were partially responsible for their own task demands. Importantly, there was a significant and strong correlation (r=0.76, p<0.001) between the time to complete the first block and the time to complete the second block, which suggests that speed is a robust individual characteristic.

2.2 Self-training

An experiment was conducted to compare three forms of feedback (no feedback, verbal feedback, and tactile feedback) while training lane keeping on a curved road. The experiment was conducted in a DDS (Figure 2). The lane width was 5 m.

Thirty male participants without any driving experience were randomly assigned to one of three conditions. One group (n=10) drove without feedback on their lane keeping performance; the students had to use task-intrinsic feedback only. A second group (n=10) was provided with verbal feedback based on their lateral deviation from the lane center (too much to the left/right when approaching the lane boundaries). A third group (n=10) received vibratory, tactile feedback from the seat bottom (Figure 2) as a function of the absolute deviation from the lane center.

A classical pretest-posttest design was employed. The experiment consisted of three subsequent sessions: a 115 s pretest (no feedback for any of the three groups), a 600 s training session on lane keeping (different feedback for the three groups), and a 115 s posttest (no feedback for any of the groups, same route as in the pretest). Speed control was automated in the entire experiment; participants only had to steer. Participants were instructed prior to the training sessions by means of written handouts to drive as properly as possible within the right lane. The task instructions also stated that the experiment comprised three sessions, the final of which was a driving test. The dependent measure was the standard deviation of lateral position (SDLP), a measure that has been shown to be a sensitive descriptor of lane keeping accuracy [9]. Here, a distinction was made between the SDLP of the whole session and a grand mean based on the SDLPs of road segments.

The data of one participant from the ‘no feedback’ group were recorded incorrectly and therefore not used in the analysis. The results (Figure 3) showed that – although there were large individual differences in pretest and posttest performance – learning had occurred for all three groups. That is, participants of all groups performed significantly better in the posttest than in the pretest (p<0.05 for all 6 comparisons using a paired t-test). When expressing the amount of learning as the pretest-posttest SDLP difference divided by the pretest SDLP, the mean learning was 35% for no feedback, 21% for verbal feedback, and 36% for tactile feedback. These numbers were not significantly different (F=2.12; p=0.14, using a one-way ANOVA). For the grand mean SDLP based on road segments, learning was 25% for no feedback, 9% for verbal feedback, and 27% for tactile feedback. These numbers were significantly different (F=3.77, p=0.037).

This experiment showed that participants were able to learn the (predominantly visual) lane keeping task in the driving simulator without explicit speech or tactile feedback on performance. Tactile feedback was effective, whereas verbal feedback tended to be less successful than the other two methods. A possible cause of the relative ineffectiveness of speech feedback could be information overload. Another of our studies showed that multimodal feedback for route instructions led to improved driver performance [10]. Theories of skill acquisition predict that feedback and instruction can be useful for enhancing declarative knowledge and directing the student’s attention to those aspects of the task that are important [11]. However, instruction and feedback can also be disruptive, because working memory limitations are exceeded, student’s attention can be misdirected, or the student can become reliant on instructions [11]. Hence, the key to a successful VDI is to determine in which situations students actually need assistance and in which situations they are fine by themselves. Once the goal of the lesson and the necessary performance criteria are clear to the student, less feedback may result in more learning.

3. THE VIRTUAL DRIVING INSTRUCTOR

3.1 Complexity or simplicity?

A seemingly evident way to improve training effectiveness is to enhance the VDI’s intelligence so that it can better understand the
Sophisticated techniques can be used, such as expert systems or in the type of intersection, or had incorrect viewing behavior [4]. It has been argued before, however, that the key to improve training effectiveness does not lie in emulating a human instructor in real time, or having a human assistant in the simulator; instead, the advantages of simulators should be better exploited [2][3]. Several caveats are in order concerning the construction of complex software. Managing both hardware and software of a training simulator is an expensive and time-consuming process [12]; a more complex computer code may be detrimental to the cost-effectiveness of a simulator. Moreover, research in the related area of on-road advanced driver assistance systems (ADAS) has shown that it is essential that the driver has a clear understanding of the system [13]; a flexible and dynamic VDI can be counter-effective, since the student may fail to grasp what the normative driving criteria are. Finally, as a general principle, simple models of human behavior are preferred over more complex models as they are more easily falsified (e.g., [14]). It is therefore better to limit the complexity and keep the software simple when possible.

3.2 Regressive instruction

As shown in Section 2.2, it is possible to learn without feedback when the environment features sufficient task-relevant information. However, for learning more complex and less visual-based tasks, such as driving away, feedback and especially instructions are considered crucial. Yet, the amount of feedback and instruction should decrease with increasing practice, since the student should eventually be able to carry out a task autonomously. Current DDSs feature automatic regressive instruction, which is a relatively simple form of student-adaptive training. When the student performs a task for the first time, he or she receives step-by-step instructions (level 1). After successfully completing the task a number of times, the student promotes to level 2, which features only corrective feedback on students’ mistakes. The third and highest level assumes that the student can act autonomously. In case that the performance of the student drops, the VDI reverts to a lower learning level [6]. To evaluate the effectiveness of regressive instruction in learning the driving away task, data were collected of all students who completed a training session in a DDS in the Netherlands in the period August 2005 – March 2006. The session started with video instructions and demonstrations, after which students had to repeatedly drive away and bring the car to a full stop for 10 minutes on a straight road.

Figure 3. Mean SDLP in the pretest and posttest. The smaller markers are depicted at the mean ± SD (n=9 for no feedback, n=10 for speech feedback, n=10 for tactile feedback). Left: SDLP of the whole session; right: grand mean based on road segments’ SDLP.

Figure 4 shows the students’ learning levels as a function of attempt. 84% of the students were able to autonomously drive away within 14 attempts, indicating that the regressive instruction was successful for this task (according to the VDI’s criteria for success). Moreover, students’ efficiency improved with practice (not shown in graph); the task completion times decreased with attempt number (the first attempt to get the car moving lasted 31.9 s on average; the fifth attempt 18.2 s, and the tenth attempt 14.6 s). Nevertheless, 16% of the students were not in level 3 at the 14th attempt. These were regularly cases in which a student repeatedly failed the task for the same reason. Independent analyses showed that certain procedural errors, such as incorrect gear selection, were not recognized by the VDI. Hence, the VDI does not recognize and remedy all types of failures that students make during this task, signifying the need for improvement.

It is recommended to conduct extensive user tests and software checks to investigate whether the strictness of the assessment criteria are appropriate, whether the instructions are non-ambiguous, whether the timing of instructions is correct, whether there are no software bugs, etcetera. In previous research, it has been suggested that such changes may lead to considerable gains in training effectiveness [15]. This is not student-adaptation as such, but rather a method to enhance overall didactic quality.

Figure 4. Number of students (percentage of sample) who performed the driving-away task at learning level 1, 2 or 3 as a function of attempt. The number of students with at least 1 attempt was 2048; the number of students with 14 or more attempts was 1520. The session lasted 10 minutes.
It is concluded that regressive instruction based on past performance – a relatively simple form of VDI adaptation – can be effective in letting students drive away autonomously. There are indications that gains in training effectiveness can be achieved by optimizing the quality of feedback and instructions (thereby addressing all students). Improving the intelligence of the VDI should be done only with careful consideration, as the associated increase in complexity may ultimately harm training effectiveness.

4. THE STUDENT-PROFILE

This section investigates how the study of individual differences can be used in constructing a student-profile. The idea is that a student-profile can be an important guide in decision making, for example about which lessons the student should follow, when to transfer to a real car, or in predicting whether someone is an accident-prone driver or not. The student-profile can be constructed from driver performance that is objectively recorded, and/or from individual characteristics (e.g., age, gender, personality).

4.1 Norm referenced assessment

Previous analysis of driver simulator training performance records has shown that task-difficulty is uneven between tasks and between sessions [16]. This means that a student’s task score does not only depend on his or her competence, but on average, it is a function of the software and strictness of the assessment criteria. Norm referenced assessment was proposed as a solution; that is, the student’s performance is transformed to percentiles relative to all students who had previously completed the same training program [16]. Figure 5 illustrates this principle based on performance records of a particular task (stopping in front of a stop sign) in a particular session. Norm-referenced assessment is a regular procedure in training and testing. The IQ test is likely the best known example (having a mean of 100 and a SD of 15 points).

As a spin-off of the normalization principle, present DDSs provide norm-referenced grades on a scale from 1 to 10 on a matrix of task-session combinations [5]. These so-called strength-weakness reports are used to assist human supervisors to decide which driving lessons should be completed by the student. An auxiliary spin-off has been the calculation of student’s mean grade on all tasks. This opportunity is currently being exploited in an online driver training competition [3][5]. At present, more than 2,500 students have participated on a voluntary basis in this online ranking to compete for prizes such as free driving lessons. One supposed key advantage is that this competitive effect improves the motivation to perform as well as possible, within all the traffic rules. A recent study from Sweden found a positive correlation between experience with computer games and skill-oriented aspects of car driving, while there were no negative effects on attitude-oriented variables [17].

A drawback of norm-referenced assessment is that the progress of a population cannot be measured. We stress that the population performance should not be used as a substitute for the norm. In the end, every driver should be able to safely carry out the crucial driving tasks. Nonetheless, norm referenced assessment allows getting an objective indication of where the student stands with respect to the population. This is impossible during training on the road where performance is not stored into a database.

4.2 Driver assessment using factor analysis

As a follow-up study of the normalization process, we investigated whether the statistical method factor analysis can be used for driver assessment [8]. Factor analysis is a technique that goes beyond the scores of individual tasks in order to reveal the underlying latent structure. To illustrate, Figure 6 shows the number of times a student had the failure ‘driving too fast’ versus the number of times the student had the failure ‘following too closely’. To prevent any causal physical relationship, ‘driving too fast’ was counted for even training sessions only and ‘following too closely’ for odd sessions. A positive correlation (r=0.58, p<0.001) exists between these two variables, indicating that they are partly redundant and possibly governed by the same common factor (i.e., a student’s tendency for violations).

Factor analysis uses the matrix of correlations amongst a great many variables to extract a small number of factors that explain the driver behavior. Using this methodology, a speed-score, error-score (or inversely: accuracy), and violation-score have been calculated for each student based on their task scores and mean task completion times [8]. The factor scores showed predictive validity with respect to the results of the on-road driving test [18]. Correcting those students with a high violation-score is particularly important for road safety as research has shown that deliberate violations (rather than errors) are predictive of road crashes [19].

4.3 Individual characteristics

This section explores the effects of age, gender, and personality characteristics on performance in driving simulators.

It is well established that young drivers have greater tendencies towards risk factors such as speeding, going out at night, sometimes in combination with alcohol or drugs [20]. Elderly drivers, on the other hand, tend to have lower physical and mental capabilities than younger drivers [21]. A number of studies have shown that, in the simulator, young experienced drivers adopt higher speeds and make fewer errors than older drivers.
these individual predictors and driver performance were always safety-margins, lane keeping accuracy). Correlations between personality traits) and performance in the simulator (speed, IQ...)

As part of previous work, we established significant correlations with driving test results [8][18]. There were no indications though that the violations (e.g., speeding, following too closely) than women (see also Figure 6). There were no indications though that the simulator was unfair; learning rates were comparable, as well as correlations with driving test results [8][18].

On the roads, men drive more than women and are more involved in risk factors such as speeding and sensation-seeking [20]. In the Netherlands in 2007, 64 men between 18-24 years of age died in car crashes as compared to 10 women (other modes of transportation than cars excluded) [25]. Women, on the other hand, are more likely to be involved in errors and accidents that are related to operational control, such as low-speed maneuvering [26]. Gender differences during simulation-based driver training were generally large, up to 1.2 standard deviations [8][18]. On average, men made fewer (steering) errors, had lower mean task completion times (a higher speed-score), and made more violations (e.g., speeding, following too closely) than women (see also Figure 6). There were no indications though that the simulator was unfair; learning rates were comparable, as well as correlations with driving test results [8][18].

As part of previous work, we established significant correlations between several personality scales (e.g., Driver Behaviour Questionnaire [27], Sensation Seeking Scale [28], and Big Five personality traits) and performance in the simulator (speed, safety-margins, lane keeping accuracy). Correlations between these individual predictors and driver performance were always lower than 0.50. This means that a certain share (<25%) of the variance can be explained by a single predictor. The most powerful predictor that we have found in the literature was an intelligence test, which predicted future flight training duration with correlation of -0.6 [29]. An overview of cognitive capacity diversity in relation to computer task performance is provided in [30].

Hence, it seems feasible that a student-profile, incorporating a combination of individual characteristics can explain a substantial share of the variance of (future) driver behavior in the simulator or on the roads. More research is needed to investigate the potential benefits of a factor-score-, gender-, age-, personality-, and/or intelligence-differentiated training program.

### 5. IMPACT ON RESEARCH & INDUSTRY

This paper provided recommendations for advancing cost-effective, student-adaptive initial driver training.

Results show that: 1) The key to a successful virtual driving instructor (VDI) is to determine in which situations the students actually need feedback and instructions. Particularly for novices, less feedback can result in better training; 2) It is recommended to be careful with respect to increasing the VDI’s complexity as this could be counter-effective. Regressive instruction based on past performance – a relatively simple form of VDI adaptation – was successful in letting students drive away autonomously. Future developments should be directed towards optimization of feedback and instructions; 3) A student-profile that incorporates a combination of individual characteristics can likely explain a substantial share of the variance in driver behavior. The remediation of deviant behavior of those with a high violation-score is particularly relevant to road safety. Further research is needed to evaluate how the training should be tailored towards these individual differences. A strength-weakness report based on the research shown in Section 4.1 is currently used in Dutch Driving Simulators.

A remark is made with respect to data storage. The present study drew heavily on the analysis of large amounts of data stored in simulators across the Netherlands. Remarkably though, in a survey of simulators (on military sites) it was found that only very few simulators had facilities for long-term data storage [12]. This makes it impossible to evaluate training effectiveness of different forms of feedback and instruction. We therefore stress the importance of data storage facilities and research for evaluating and improving the effectiveness of a driver training system.

### 6. ACKNOWLEDGMENTS

The research reported in this paper has been funded by the Dutch Ministry of Economic Affairs, under its Innovation Program on Man-Machine Interaction, IOP MMI. We thank the student team that contributed to the experiment. We express thanks to Green Dino Virtual Realities for their data storage facilities.

### 7. REFERENCES


