

Violations and errors during simulation-based driver training

de Winter, JCF; Wieringa, PA; Mulder, JA; Mulder, M; Kuijpers, J.

Publication date

2007

Document Version

Accepted author manuscript

Published in

Ergonomics: an international journal of research and practice in human factors and ergonomics

Citation (APA)

de Winter, JCF., Wieringa, PA., Mulder, JA., Mulder, M., & Kuijpers, J. (2007). Violations and errors during simulation-based driver training. *Ergonomics: an international journal of research and practice in human factors and ergonomics*, 50(1), 138-158.

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

Violations and errors during simulation-based driver training

J. C. F. de Winter, P.A. Wieringa, J. Kuipers, J. A. Mulder, and M. Mulder

De Winter, J.C.F., Wieringa, P.A., Kuipers, J., Mulder, J.A., & Mulder, M. (2007). Violations and errors during simulation-based driver training. *Ergonomics*. 50, 138–158. <https://doi.org/10.1080/00140130601032721>

Delft University of Technology, j.c.f.dewinter@tudelft.nl

Abstract

The effectiveness of virtual driving instruction can increase when techniques that automatically distinguish between violations and errors are available, two behaviours requiring different types of remediation. This study reports the analysis of the objectively measured performance of 520 participants completing a simulation-based training programme. Factor analysis of failure reasons showed that violations and errors were the primary underlying factors. Men committed more violations and women made more errors; the magnitude of gender differences corresponded to the factor loadings. Factor analysis of the mean task completion times yielded a factor that can be described as the extent to which motivation for speed was expressed in quicker task execution. Quicker participants completed more tasks, committed more violations but made fewer errors. Participants reduced errors during forced-paced driving and increased speed during self-paced driving. The authors recommend exploiting the distinction between violations and errors by developing interfaces and feedback for both types of aberration.

1. Introduction

In these days of high-priced fuel, simulators provide a cost-effective solution to initial driver training. Besides financial benefits, simulators offer great opportunities for carrying out objective measurements on the user's actions in a safe and purpose-developed virtual environment (Vlakveld, 2005). Not surprisingly, many driving schools have started using simulators for training their students. It has been estimated that around 100 low-cost simulators are currently used in the Netherlands for initial driver training (Kappé & Van Emmerik, 2005). Nowadays, the introduction of virtual instructor software, sometimes referred to as an intelligent tutoring system, contributes to further automation of driver education, the result of which changes the role of the human instructor into that of a supervisor (Kappé & Van Emmerik, 2005; Weevers et al., 2003a, 2003b).

In comparison with a human instructor, a virtual instructor monitors and assesses the driver's actions by comparing his or her performance with normative performance and gives corrective feedback when the driver fails to comply with these norms (e.g., Michon, 1993). However, there are indications that failures cannot always be effectively remedied by practice and simple feedback on performance. It has been found that learner drivers may prefer to increase speed rather than to drive more accurately to comply with the

norms (De Winter et al., 2006). This distinction between speed and accuracy seems to be similar to the distinction between violations and errors, defined here as intentional and unintentional deviations from normative performance respectively (see also Rothengatter, 1997). It has been suggested that errors could be counteracted by means of proper training (e.g., Reason et al., 1990), whereas violations could be better prevented by self-reflection and attitude enforcement (e.g., Hatakka et al., 2002). Because both types of behaviour require different modes of communication, a virtual instructor should be able to distinguish between intentional violations and unintentional errors. Therefore, objective methods are needed to make a distinction.

The violation-error distinction has been extensively studied by means of questionnaires. Reason et al. (1990) were the first to show, using the Driver Behaviour Questionnaire (DBQ), that driving behaviour on the roads can be divided into violation and error components. They suggested that violations and errors are caused by different psychological processes. Others have repeated and extended this work. For example, demographic differences and the relation of violations and errors to road accidents have been investigated (Åberg & Rimmö, 1998; Blockey & Hartley, 1995; Kontogiannis et al., 2002; Lajunen et al., 2004; Mesken et al., 2002; Özkan & Lajunen, 2005; Parker et al., 1998; Rimmö & Åberg, 1999; Xie & Parker, 2002).

Although violations and errors are probably governed by different psychological processes, this does not imply that both types of aberration are unrelated. There are numerous indications that drivers show adaptive behaviour. For example, it has been reported that skid control training indeed improves vehicle-handling skills but that it also leads to an increased number of speed violations (Katila et al., 1996), and Van Winsum and Godthelp (1996) showed that drivers with better steering skills adopt higher speeds in bends. Furthermore, speed control plays an important role in driving behaviour because speed influences task conditions and hence the susceptibility to violations or errors (Rothengatter, 1997; Fuller, 2005). Without doubt, violations, errors, and speed are intimately related and further investigation is therefore needed to support the development of virtual instructors.

Distinguishing between violations and errors without completing questionnaires is easier said than done. First, a computer cannot assess a person's intentions directly, although progress is made in correlating physiological measurements with driving performance (e.g., Lin et al., 2005). Second, there exists no acceptable definition of normative driving behaviour (Kappé & Van Emmerik, 2005; Pirenne et al., 2002; Rothengatter, 1997). Traffic law regulations do not suffice, because they implicitly assume intentional behaviour. Driving through a red light, for example, is always regarded as a violation, even if the driver erroneously failed to see the light. Third, during an action sequence, violations and errors can occur simultaneously (Reason et al., 1990) and therefore observed driving behaviour can be the result of both a violation and an error.

These difficulties also play a role in the development of driver support systems (Rothengatter, 1997). Rothengatter (1991) proposed to simplify the problem by attributing every deviation from normative performance to intentional violations. Another option is "to consider deviations from normative performance as intentional only when these 'normally' occur with specific drivers" (Rothengatter, 1997). In other words, given the fact that a group of violation-prone drivers (for example, men) is observed to frequently devi-

ate from particular normative performance (for example, not keeping enough distance), this deviation can generally be considered a violation. Similarly, Reason et al. (1990) stated that: “Distinctions between errors and violations, on the one hand, and their respective contributions to road accidents, on the other, emerge more clearly from group analyses in which age and sex differences, rather than individuals, are the focus of study” (p. 1317). Gender differences have been identified in DBQ scores, with men reporting more violations than women and women reporting more errors than men (Åberg & Rimmö, 1998; Blockley & Hartley, 1995; Parker et al., 1995; Reason et al., 1990).

The aim of this study was to find out whether violations and errors could be identified in objective performance data. For this purpose, performance data of participants who completed a simulation-based driver training programme were analysed and the relationships between violations, errors, and speed were assessed.



Figure 1. Dutch Driving Simulator (Green Dino, 2007).

2. Method

For a period of 8 months, performance data were recorded for participants involved in simulation-based driver training in simulators placed throughout the Netherlands.

2.1. Hardware

The driving simulators used in this study were the commercially available fixed-base Dutch Driving Simulators (DDS; Green Dino, 2007) (Figure 1). The vehicle controls of the DDS were similar to an actual car with a manual transmission. Force feedback was provided on the steering wheel according to the self-aligning torque of the front wheels and acceleration cues were provided by means of vibrations on the steering wheel column and vibrations in the back of the seat. Three projectors provided a geometrical 180° field of view. The resolutions were 1024 x 768 pixels for the front view projection and 800 x 600 pixels for the side view projections. The dashboard, the vehicle interior, and mirrors

were integrated in the projected image.

2.2. Training software

The curriculum was based on Dutch driver training and consisted of 15 lessons, 27 min each. Lessons 1–5 were dedicated to vehicle control, lessons 6–10 to driving in urban areas with intersections and roundabouts, and lessons 11–15 to motorway driving. A human supervisor was able to manually alter the default order of the lessons. Each lesson consisted of two or three sessions with the duration ranging from 4 min 30 s to 16 min 4 s. Sessions were preceded by instructive text, voice, and movies. In total, the training programme consisted of 1 h of such multimedia instructions and 5 h 45 min of simulation-based driving.

A virtual instructor provided instructions and feedback on performance. After the task was successfully executed a successive number of times, instructions and feedback were reduced. Tasks were automatically activated and deactivated according to predefined conditions. For example, the task of taking bends was activated when entering the bend and deactivated when leaving the bend. Tasks could also end as a result of a 30-s time-out. Different tasks could be activated within a session, but only one task could be activated at a time. Tasks were automatically assessed by a virtual instructor that monitored whether the participant exceeded predefined (normative) thresholds. Most tasks could be failed for different reasons. More information about the virtual instructor can be found in Weevers et al. (2003a, 2003b).

2.3. Participants

A total of 42 different driving schools (totalling 46 DDSs) across the Netherlands took part in the measurements. Performance data of 2,530 participants were recorded, 520 of whom had completed all lessons in default order (263 men, 257 women, mean age based on available records =19.7 years). As these participants had completed the same training programme, their data were used for the analysis that is presented in this chapter. The median number of days needed to finish the default training programme was 17.5. After having completed the simulation-based training programme, participants continued training in a real vehicle.

2.4. Measures

Driving performance on 20 tasks (see Table 1) was automatically recorded. All data were brought together after completion of the 8-month measurement. The number of successes, the number of failures per failure reason, and the mean task completion time (MTCT) were calculated for every participant for every task. If a participant had failed a task for more than one reason, then only the failure reason listed first in Table 1 was recorded. All time-outs were excluded from the analysis. Only task successes were used to calculate MTCT; the MTCT was not calculated when the participant was never successful on a task. The number of task activations for a participant-task combination was defined as the number of successes plus the number of failures.

Table 1. Means and standard deviations of the number of successes for each task (T1 to T20), the mean task completion time (MTCT) for each task, and the number of failures for each failure reason (F1 to F52) (Note that failure reasons belong to task)

Task	Successes		MTCT		Failure reason	Failures	
	Mean	SD	Mean	SD		Mean	SD
T1 Driving away during start/stop exercise ^a	12.79	4.64	15.32	4.95	F1 Stalling the engine	0.06	0.34
					F2 Throttle > 75% when starting engine	0.09	0.38
					F3 Wrong pedal or wrong gear	0.17	0.47
					F4 Clutch pedal speed > 100% per s	2.20	2.26
					F5 Throttle > 75% when in gear	0.12	0.48
T2 Driving away from parking lane	1.42	1.36	31.01	11.76	F6 Stalling the engine	1.54	2.03
					F7 Wrong pedal or wrong gear	0.17	0.46
					F8 No indicator during lane change	0.77	1.07
					F9 Lateral speed > 0.5 lane per s	0.09	0.35
T3 Driving on and near intersections	82.35	11.57	38.63	3.66	F10 Collision with other car, pedestrian, or bicyclist	0.98	1.13
					F11 Driving too fast with respect to others having right of way	16.88	3.98
					F12 Driving too fast (no need to stop)	12.62	7.45
					F13 Approaching red traffic light too fast	1.50	1.54
					F14 Pressing clutch when decelerating and having to turn	6.22	4.81
					F15 Not in 2nd gear when having to turn or near other car(s)	2.30	3.57
F16 Approaching amber traffic light too fast	1.95	1.37					

(table continues)

Table 1. (continued)

Task	Successes		MTCT		Failure reason	Failures			
	Mean	SD	Mean	SD		Mean	SD		
T4	Keeping maximum speed in bends	17.06	4.33	17.31	1.98	F17	Driving more than 5 km/h too fast	3.46	3.18
T5	Keeping maximum speed on motorway segments	52.42	10.55	20.13	6.86	F18	Driving more than 5 km/h too fast	16.56	7.98
T6	Keeping safe distance in intersections world	0.59	1.04	10.36	3.01	F19	Time headway < 1 s	0.22	0.60
T7	Keeping safe distance on winding road	1.69	2.73	11.13	6.39	F20	Time headway < 1 s	0.90	2.07
T8	Keeping safe distance on motorways	10.85	6.49	14.00	3.21	F21	Time headway < 1 s	8.34	6.78
T9	Lateral motorway manoeuvres	171.47	31.53	15.84	2.23	F22	No indicator at exit lane or during lane change	27.54	19.74
						F23	Wrong pedal or wrong gear	0.14	0.43
						F24	Lateral speed greater than half a lane per s	4.72	6.15
T10	Lane tracking ^b	4.10	2.08	29.04	1.96	F25	Lane centre error > 1.40 m	1.49	1.43
						F26	Lane centre error more than 2 s between 0.85 m and 1.40 m	0.62	0.92
						F27	Pressing clutch when decelerating in turn	0.55	0.88
						F28	Lane centre error less than 2 s between 0.85 m and 1.40 m	2.09	1.47

(table continues)

Table 1. (continued)

Task	Successes		MTCT		Failure reason	Failures	
	Mean	SD	Mean	SD		Mean	SD
T11 Shifting gears	53.33	13.20	9.81	4.26	F29 Pressing wrong pedal/shifting to wrong gear	0.06	0.29
					F30 Clutch pedal speed > 200% per s	13.94	9.83
					F31 Throttle > 75% when in gear	0.83	1.60
T12 Shifting gears during shifting gears exercise a,b	13.11	4.28	9.70	3.11	F32 Wrong pedal or wrong gear	0.59	0.82
					F33 Clutch pedal speed > 200% per s	0.65	1.17
					F34 Throttle position > 75% when in gear	0.16	0.50
T13 Speed tracking a,c,d	10.61	2.50	17.73	0.32	F35 Deviating more than 5 km/h from instructed speed	1.60	1.75
T14 Spotting objects (spatial-perceptual task) a,b,d	18.14	1.57	18.83	0.88	F36 Not pressing horn before passing the object (too late)	0.90	1.12
					F37 Pressing horn while object is not yet visible	0.40	0.77
T15 Stopping during start/stop	16.67	4.30	15.84	6.23	F38 Stalling the engine	0.47	2.10
T16 Stopping in front of stop	2.49	1.69	39.40	11.44	F39 Stalling the engine	0.16	0.41
					F40 Brake pedal speed > 1500% per s	0.11	0.34
					F41 Pressing clutch when engine speed > 1200 rpm	1.45	1.30
					F42 Not pressing clutch so that engine almost stalls	1.62	1.58

(table continues)

Table 1. (continued)

Task	Successes		MTCT		Failure reason	Failures		
	Mean	SD	Mean	SD		Mean	SD	
T17 Stopping the car in parking lane	3.45	1.51	40.21	10.38	F43	Driving on soft shoulder	0.01	0.11
					F44	No indicator during lane change	0.76	0.98
					F45	Lateral speed > 0.5 lane per s	0.15	0.52
T18 Taking bends during exercise	5.64	2.57	10.39	3.83	F46	Lane centre error > 1.0 m and increasing more than 0.3 s	3.07	2.21
					F47	Driving faster than 30/60 km/h in sharp/normal turns	2.08	1.58
					F48	Pressing clutch when decelerating in turn	1.18	1.48
T19 Taking bends during exercise ^b	5.94	2.81	7.02	1.91	F49	Lane centre error > 1.0 m and increasing more than 0.3 s	5.68	2.79
T20 Taking exit and entry ramps	21.73	11.23	17.71	5.30	F50	Lane centre error > 1.0 m and increasing more than 0.3 s	19.05	8.56
					F51	Driving faster than 30/60 km/h in sharp/normal turns	5.37	3.76
					F52	Pressing clutch when decelerating in turn	0.73	1.11

^a Automated steering

^b Automated speed of the car

^c Reactivated at fixed times

^d Automated gear changing

3. Results

3.1. Frequencies

Table 1 shows the means and standard deviations of the number of failures per failure reason, the number of successes per task, and the MTCT per task. It can be seen that the mean number of successes and mean number of failure reasons differed highly between tasks, which can be explained by the curriculum design and the strictness of the normative performance. For example, lateral highway manoeuvres (T9) was activated many times, which can be explained by the fact that multiple sessions were devoted to lane changes and overtaking. Stopping in front of a stop sign (T16) is an example of a task that was activated less frequently, because it was given in one short-lasting session only and the mean MTCT was relatively high. Because the assessment criteria of this task were rather strict, on average, more failures (3.34 which is the sum of failure reasons F39, F40, F41, and F42) than successes (2.49) were counted.

An interesting finding was that the average man had more failures than the average woman (184 vs. 167, Cohen's d effect size (d) = 0.37, $p < 0.001$ using a t test) and more successes as well (521 vs. 491, $d = 0.56$, $p < 0.001$). In other words, men were involved in more task activations, even though every participant received equal training time.

3.2. Factor analysis of failure reasons

3.2.1. Factor analysis

To investigate whether a distinction can be made between violations and errors, the Pearson correlation matrix among the failure reasons was constructed and factors were extracted using the maximum likelihood method. Only two factors were extracted, because the violation-error distinction was of interest. The scree plot clearly supported this two-factor solution. Parallel analysis (O'Connor, 2000) indicated that an eight-factor solution would be most appropriate. However, the eight-factor solution was not readily interpretable and was clearly an overextraction resulting from the fact that parallel analysis is sensitive to sample size thereby recommending an inappropriately high number of factors. Oblique rotation (oblimin) was tried and it was found that only a small negative correlation (-0.10) existed between the factors, so it was decided to simplify the solution by orthogonal Varimax rotation. The result of the factor analysis is shown in Table 2. The two orthogonal factors accounted for 11.9% of the total variance: 7.35% for factor 1 and 4.52% for factor 2. The small variance can be explained by several failure reasons that hardly contributed to the overall solution. After removing 34 failure reasons having a correlation with all other failure reasons that was smaller than 0.30 (Tabachnick & Fidell, 2001), the two-factor solution would explain 25.3% of the variance, which is comparable to the DBQ-study of Blockey and Hartley (1995). Failure reasons loading above 0.20 were considered high enough (salient) to assume that a relationship existed between the failure reason and the factor (Gorsuch, 1974, taking into account the sample size of 520).

Table 2. Factor loadings and communalities (h^2) of failure reasons in the violation-error solution

Failure reason	Violations	Errors	h^2
F1 Stalling the engine	0.06	0.09	0.01
F2 Throttle > 75% when starting engine	0.00	0.03	0.00
F3 Wrong pedal or wrong gear	0.02	0.23	0.05
F4 Clutch pedal speed > 100% per s	0.28	0.04	0.08
F5 Throttle > 75% when in gear	0.00	0.08	0.01
F6 Stalling the engine	0.01	0.20	0.04
F7 Wrong pedal or wrong gear	-0.06	0.22	0.05
F8 No indicator during lane change	0.11	0.02	0.01
F9 Lateral speed > 0.5 lane per s	-0.01	0.14	0.02
F10 Collision with other car, pedestrian, or bicyclist	0.13	0.04	0.02
F11 Driving too fast with respect to others having right of way	0.29	-0.08	0.09
F12 Driving too fast (no need to stop)	0.43	0.28	0.26
F13 Approaching red traffic light too fast	0.09	0.24	0.07
F14 Pressing clutch when decelerating and having to turn	0.39	-0.25	0.21
F15 Not in second gear when having to turn or near other car(s)	0.07	0.16	0.03
F16 Approaching red traffic light too fast	-0.14	0.17	0.05
F17 Driving more than 5 km/h too fast	0.57	-0.17	0.35
F18 Driving more than 5 km/h too fast	0.72	0.06	0.53
F19 Time headway < 1 s	0.46	0.18	0.25
F20 Time headway < 1 s	0.44	-0.03	0.19
F21 Time headway < 1 s	0.62	-0.01	0.39
F22 No indicator at exit lane or during lane change	0.27	0.07	0.08
F23 Wrong pedal or wrong gear	-0.01	0.11	0.01
F24 Lateral speed greater than half a lane per s	0.19	0.29	0.12
F25 Lane centre error > 1.40 m	0.01	0.45	0.20
F26 Lane centre error more than 2 s between 0.85 m and 1.40 m	-0.10	0.32	0.11
F27 Pressing clutch when decelerating in turn	0.29	-0.46	0.30
F28 Lane centre error less than 2 s between 0.85 m and 1.40 m	0.06	0.21	0.05
F29 Pressing wrong pedal/shifting to wrong gear	0.00	0.06	0.00
F30 Clutch pedal speed > 200% per s	0.68	-0.09	0.47
F31 Throttle > 75% when in gear	0.25	0.09	0.07
F32 Wrong pedal or wrong gear	-0.12	0.23	0.07
F33 Clutch pedal speed > 200% per s	0.37	-0.09	0.15
F34 Throttle position > 75% when in gear	0.21	-0.05	0.05
F35 Deviating more than 5 km/h from instructed speed	-0.05	0.32	0.10
F36 Not pressing horn before passing the object (too late)	-0.04	0.07	0.01
F37 Pressing horn while object is not yet visible (too early)	0.04	-0.01	0.00
F38 Stalling the engine	0.06	0.22	0.05

(table continues)

Table 2. (continued)

Failure reason	Violations	Errors	h^2
F39 Stalling the engine	-0.04	0.12	0.02
F40 Brake pedal speed > 1500% per s	0.13	-0.05	0.02
F41 Pressing clutch when engine speed > 1200 rpm	0.39	-0.17	0.18
F42 Not pressing clutch so that engine almost stalls	-0.29	0.34	0.20
F43 Driving on soft shoulder	-0.05	0.09	0.01
F44 No indicator during lane change	0.04	-0.01	0.00
F45 Lateral speed > 0.5 lane per s	0.27	-0.05	0.07
F46 Lane centre error > 1.0 m and increasing more than 0.3 s	-0.09	0.52	0.28
F47 Driving faster than 30/60 km/h in sharp/normal turns	0.18	0.05	0.04
F48 Pressing clutch when decelerating in turn	0.38	-0.31	0.24
F49 Lane centre error > 1.0 m and increasing more than 0.3 s	-0.13	0.59	0.37
F50 Lane centre error > 1.0 m and increasing more than 0.3 s	0.05	0.24	0.06
F51 Driving faster than 30/60 km/h in sharp/normal turns	0.36	-0.01	0.13
F52 Pressing clutch when decelerating in turn	0.05	-0.01	0.00
Eigenvalues and sum of squares	4.76	2.97	6.17
Percentage explained	7.35	4.52	11.87

3.2.2. Reliability analysis

The internal consistencies (Cronbach's α) of salient variables were 0.79 for factor 1 and 0.65 for factor 2, which seems to be fairly low but around the same level as those found after factor analysing DBQ-responses (Lajunen et al., 2004).¹ Additionally, the reliability of the factor loadings was assessed by randomly dividing the 520 participants into two groups of 260 participants and re-performing the factor analysis for these two subsamples. Factorial agreement was then assessed by the congruence coefficient (Levine, 1977). The factor loading congruence coefficients between the subsamples and the total sample were 0.94 and 0.98 and the congruence coefficient between both subsamples was 0.87, indicating that reasonably equal results could be produced by two independent (and smaller) groups of participants.

3.2.3. Factor 1 description

Table 2 shows that failure reasons loading (> 0.20) on factor 1 were almost exclusively excess of speed and the breaching of safety-related thresholds. The failure reasons loading most highly (> 0.40) on this factor were: driving too fast (F12, F17, F18); moving the clutch too quickly (F30); and following too closely (F19, F20, F21). Not using the traffic indicator (F22) and pressing the clutch when it was forbidden by the rules (F14, F27, F41, F48) also loaded on the first factor. The latter procedure followed Dutch training regula-

¹ Note that a high Cronbach's α is not a necessary condition for a good factor analysis (Boyle, 1991; appendix B). Instead, maximizing Cronbach's α can result in overly specific factors with low validity.

tions, which state that drivers should be in the desired gear before executing a manoeuvre to prevent excessive increase of task difficulty (Vissers, 2001).

3.2.4. Factor 2 description

Failure reasons that loaded (> 0.20) on factor 2 were related to poor vehicle control. The highest factor 2 loadings (> 0.30) represented lane-tracking or speed-tracking errors (F25, F26, F35, F46, F49). Lower loadings generally related to procedural failures such as stalling the engine and shifting to the wrong gear or applying the wrong pedal. Interestingly, approaching a red or amber traffic light too fast (F13, F16) and driving too fast on and near intersections (F12) loaded moderately on factor 2 as well, although the characteristics of these failure reasons resemble the characteristics of factor 1. Personal observations can explain the loadings on factor 2. Traffic lights and speed signs were relatively difficult to perceive (see also Figure 1) and often disappeared behind the projected rear view mirror, so these aberrations can be explained as the result of a perceptual deficiency in the driving simulator.

3.2.5. Negative factor loadings

Negative factor loadings have a distinct meaning as well. While positive factor 1 loadings were found for speeding and breaching safety-related thresholds, negative factor 1 loadings may represent a reluctance to act in time. Not pressing the clutch so that the engine almost stalls (F42) and approaching an amber traffic light too fast (F16) both belong to this category. While positive factor 2 loadings were related to poor vehicle control, negative loadings appear to be a sign of good vehicle control and especially occurred for applying the clutch when against the rules (F14, F27, F41, F48). It can be reasoned that participants with better vehicle-handling abilities were involved in these failures because they were less obstructed by increased task difficulty. For example, the Pearson correlation between F27 and F49 is -0.29 ($p < 0.001$) indicating that participants who had fewer lane-tracking failures more often applied the clutch in a bend.

3.2.6. Comparison to DBQ and gender differences

When comparing the factor distinction to the violation-error distinction from factor analyses of DBQ responses, it appeared that both categories are quite similar. Behaviours such as driving too close and speeding were clustered into the violation category, whereas perceptual, judgemental, and vehicle control-related deficiencies were clustered into the error categories (e.g., Reason et al., 1990).

This study supports the finding that the two factors represent violations and errors. Studies using a DBQ and a factor analysis have shown that men are generally more involved in violations and that women are more involved in errors (e.g., Reason et al., 1990). Figure 2 shows gender differences in terms of effect size (Cohen's d) versus factor loading differences (factor 1 loading minus factor 2 loading) for each of the 52 failure reasons. It can be seen that a clear relationship exists. When factor loading difference is high (failure reason is more a violation than an error), men had more failures; when the factor loading difference is low (failure reason is more an error than a violation), women had more failures. A Pearson correlation between gender differences and factor loading differences supported this observation ($r = 0.97$, $p < 0.001$). Pressing the horn too late on the spatial-

perceptual task (F36) occurred significantly more amongst women, which is in accordance with literature about gender differences in spatial abilities (e.g., Coluccia & Louse, 2004). The correspondences between factor loadings and gender differences supports that factor 1 represents violations and factor 2 represents errors.

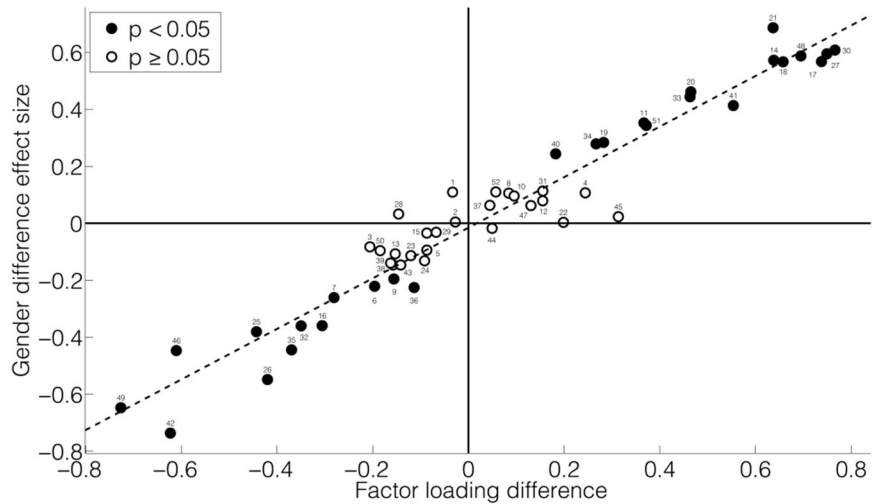


Figure 2. Magnitude of gender difference (Cohen's d effect size) of number of failures versus factor loading difference (violation factor loading minus error factor loading) for each of the 52 failure reasons.

3.3. Factor analysis of mean task completion times

3.3.1. Factor analysis

To investigate the speed choice of participants, a Pearson correlation matrix of the MTCTs of the 20 tasks was calculated; missing items were excluded pairwise. Parallel analysis indicated that two factors should be extracted; however, the two-factor solution was not regarded as interpretive. The scree plot clearly indicated that a one-factor solution was most appropriate, so it was decided to extract only one factor using the maximum likelihood method. The factor accounted for 15.6% of the total variance. The factor loadings are shown in Table 3.

3.3.2. Reliability analysis

The internal consistency (Cronbach's α) of the salient variables was 0.78. The factor loading congruence between two random subsamples of 260 participants was 0.96 and the factor congruences between each subsample and the total sample were 0.99 each, indicating that the factor loadings can be considered reliable.

3.3.3. Factor description

The extracted factor can best be described as the extent to which motivation for speed was

expressed in the MTCT and is referred to as the paceability ² of the driving task. The highest factor loadings (> 0.50) were found for tasks for which the participant was relatively unconstrained concerning the task completion time. For example, it can be imagined that the time taken to complete straight road segments (T5) or to complete lateral highway manoeuvres (T9) depended largely on the participant's preferred speed. Taking bends (T4, T18, T20), stopping the car (T15, T16, T17), and driving away from a parking lane (T2) had lower factor loadings, because the MTCT was more ambiguously related to motivation for speed. External factors, such as the task difficulty, relation to initial speed or the predefined criteria that started and ended the tasks, played a more important role here, so that the MTCT was less informative about the participant's preferred speed. Tasks for which the MTCT or the speed were computer-controlled (forced-paced tasks T10, T13, T19) had near-zero factor loadings. The distance-keeping tasks (T6, T7, T8) showed negative factor loadings, because these tasks were active when driving within 2 s distance of another car. Generally, drivers with a tendency to be quick followed the car in front more closely so that the task remained active for a longer time, so, paradoxically, participants who wanted to be quick took longer to complete distance-keeping tasks.

3.3.4. Comparison to gender differences

Again, the extracted factor was compared with gender differences. Figure 3 shows that a clear relationship exists between the paceability factor and the effect size of the MTCT of men and women. The larger the factor loading, the lower the effect size ($r = -0.96$, $p < 0.001$), indicating that men have a higher inclination for speed than women.

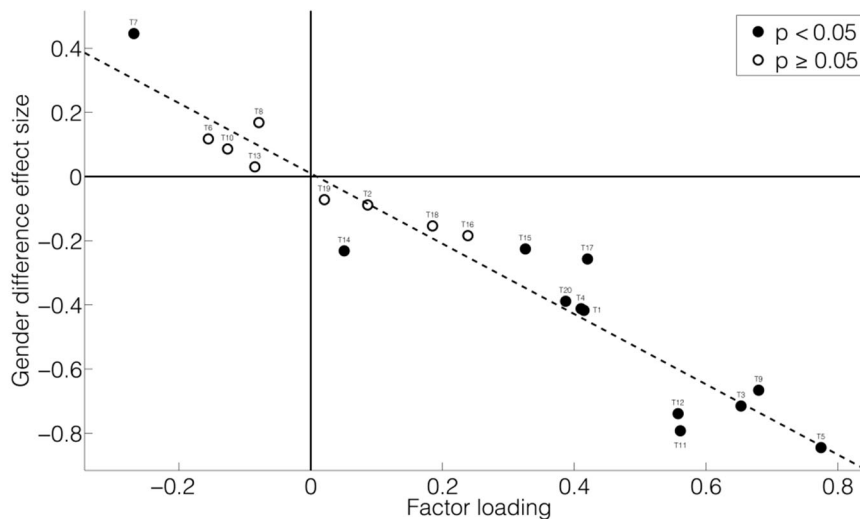


Figure 3. Magnitude of gender difference (Cohen's *d* effect size) of mean task completion time versus paceability factor-loading for each of the 20 tasks.

² In other chapters, the factor extracted from the task completion times is referred to as the "speed factor"

3.4. Factor score correlations

Violation-scores, error-scores, and speed-scores were calculated for each participant according to the Bartlett weighted least squares procedure. The sign of the speed-score was reversed so that increased motivation for speed corresponded to an increase in speed-score. Additionally, activations-scores were calculated from the mean z-transformed number of task activations; the activation-scores are an indicator for the number of tasks completed by the participant.

Figure 4 shows the activations-score versus the speed-score for each participant. It can be seen that participants having a high motivation for speed were involved in more tasks ($r = 0.78, p < 0.001$). On average, men had a higher speed-score than women (0.46 vs. -0.48, $d = 1.06, p < 0.001$) and a higher activations-score as well (0.36 vs. -0.37, $d = 0.78, p < 0.001$). The latter result corresponds to the aforementioned result that men had both more successes and more failures than women (section 3.1).

Figure 5 shows the violation-score versus the speed-score. It can be seen that speed and violations bear a strong relationship ($r = 0.71, p < 0.001$) insofar as quicker participants committed more violations. This makes sense when considering that participants with a higher speed-score were involved in more tasks (see Figure 4) and that many violations were found to be speed related. It can also be seen that participants with a very high speed-score (> 1.50) were involved in an increased number of violations. On average, men had a higher violation-score than women (0.38 vs. -0.39, $d = 0.83, p < 0.001$).

Figure 6 shows that motivation for speed had a small inverse correlation with errors ($r = -0.24, p < 0.001$). Men had a lower mean error-score than women (-0.37 vs. 0.38, $d = -0.81, p < 0.001$).

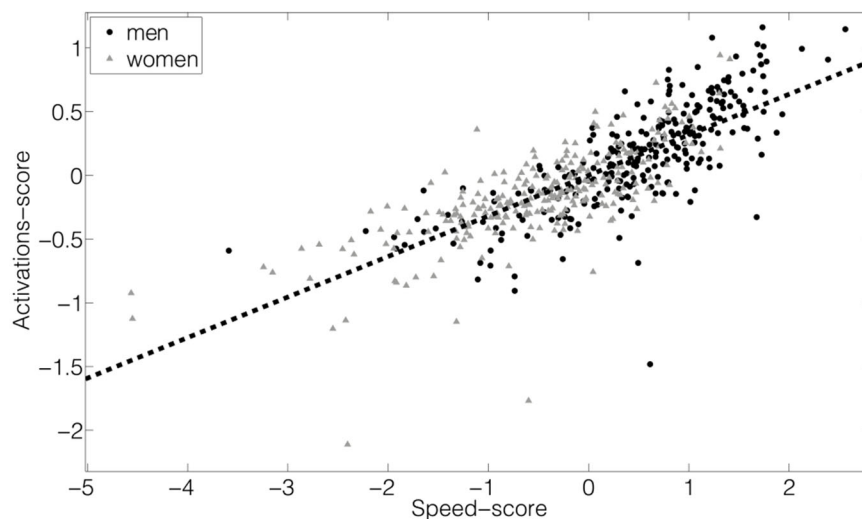


Figure 4. Activations-score versus speed-score for each of the 520 participants. The dashed line is a linear fit.

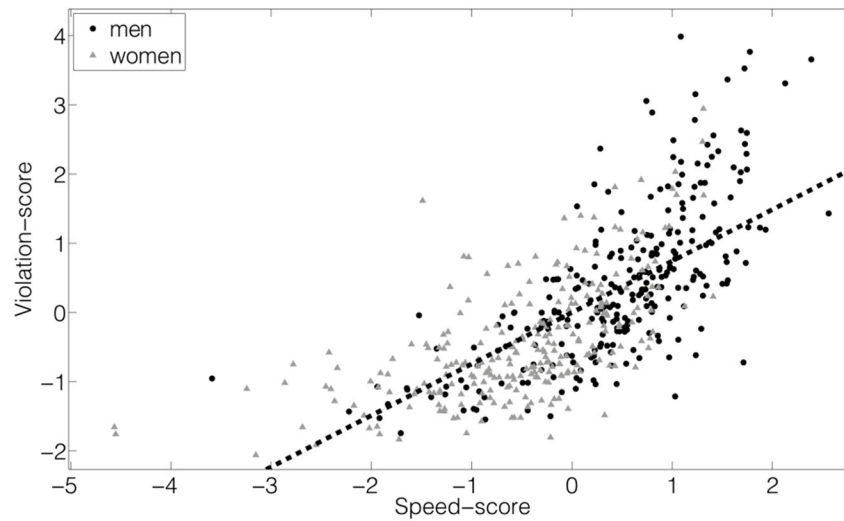


Figure 5. Violation-score versus speed-score for each of the 520 participants. The dashed line is a linear fit.

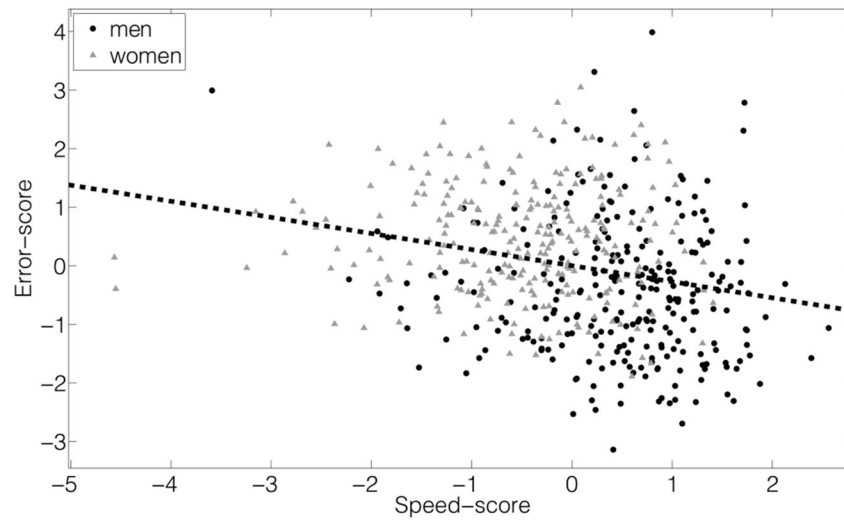


Figure 6. Error-score versus speed-score for each of the 520 participants. The dashed line is a linear fit.

3.5. Change of behaviour with time

To further investigate the relationships between errors, violations and speed, the participant's temporal behaviour within sessions was addressed. A distinction was made between forced-paced (paceability near zero) and self-paced tasks (paceability considerably greater than zero).

3.5.1. Forced-paced

For forced-paced tasks, plotting the mean cumulative number of failures and task activations versus time typically yields a graph as shown in Figure 7. This figure presents the task lateral position tracking (T10), during which participants had to drive accurately in the middle of the right lane of a winding road. The waving pattern occurred because the task was reactivated every 30 s. The solid lines represent the mean cumulative number of two failure reasons (F25 + F26). Only these two failure reasons were shown because they could be characterized as errors (see Table 2). It can be seen that the slope of these curves decreases with time, which can be explained by performance improvement. The dotted lines show that the mean number of errors can well be described by means of a simple experience curve as defined in equations (1) to (3).

$$f(t) = F0 \cdot t^b \quad (1)$$

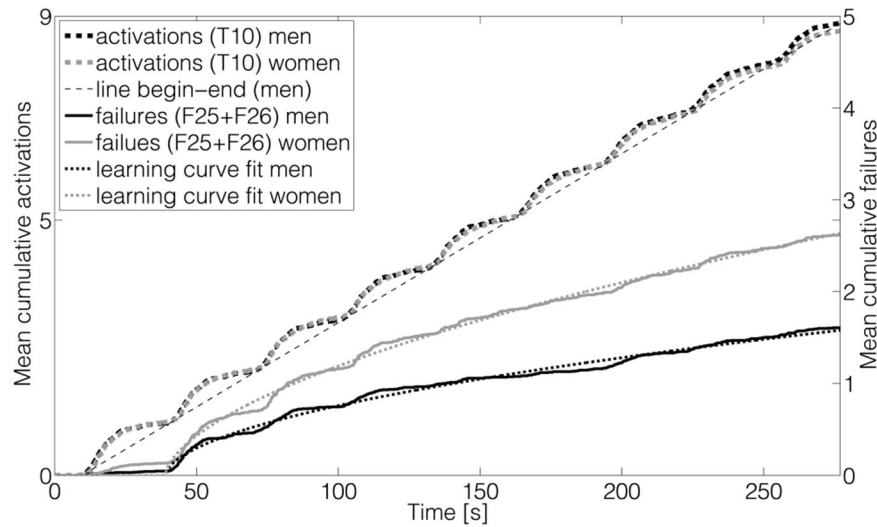


Figure 7. Mean cumulative number of task activations (task T10) and mean cumulative number of failures (failure reasons F25 and F26) versus time in the session. The thin black line through begin and endpoint was included to illustrate that the slopes of the curves remained constant with time.

Integration to obtain cumulative values yields:

$$F(t) = \frac{F0 \cdot t^{b+1}}{b+1} \quad (2)$$

In which t is the time since a starting moment, $f(t)$ is the failures per s at time t , $F0$ is the failures per s at $t = 0$, $F(t)$ is the cumulative number of failures at time t , and b is the learning constant.

The learning percentage is defined as:

$$\theta = 2^b \quad (3)$$

Best fits for men and women were obtained with reasonably similar learning percentages (θ) of 72.2 and 74.6 and with different begin values ($F0$) of 0.048 and 0.065 respectively. This indicates that men and women did not markedly differ in their learning abilities; however, they started at different initial performance levels. These similarities and differences respectively were also observed in other forced-paced tasks.

3.5.2. Self-paced

Figure 8 shows a similar graph but for the self-paced task driving away during start/ stop exercise (T1). During the session, participants had to drive away repeatedly and bring the car to a stop for 10 min on an endless straight road. No other vehicles were present and participants did not have to steer. It can be seen that the slopes of the dashed grey and black curves increase, which can be explained by the fact that participants speeded up, hence increasing the task activation rate. Looking at the mean cumulative number of violations (F4, moving the clutch too quickly), it can be seen that the slopes of these lines also increase, indicating that (simulation-based) training actually led to an increased number of violations per time unit. Similar patterns were observed in other self-paced tasks.

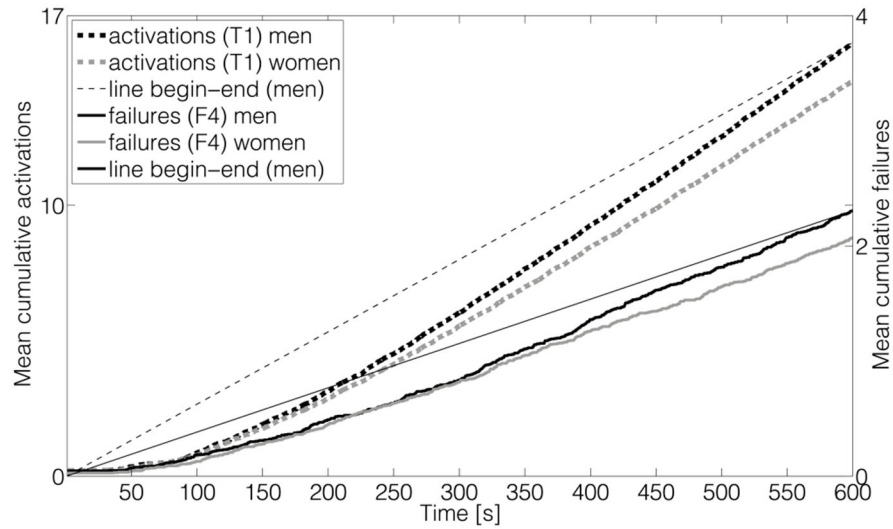


Figure 8. Mean cumulative number of task activations (task T1) and mean cumulative number of failures (failure reason F4) versus time in the session. The thin black lines through begin and end-points were included to illustrate that the slopes of the curves increased with time.

4. Discussion

4.1. Distinction between violations and errors

Two factors were identified in the automatic recordings of failures in participants who completed a simulation-based driver training programme. This two-factor solution appeared to be similar to the violation-error distinction obtained from questionnaires, such as the DBQ, that asked respondents about intentional and unintentional driving behaviours. Although the main criterion of distinguishing between violations and errors is intention, which was not measured in this study, the results support the finding that the extracted factors represented violations and errors. Factor loadings indicated that differences exist in the extent to which a failure can be regarded violation and error, and gender differences corresponded with these differences. Men committed more violations and women made more errors; the differences being the largest for the failure reasons that loaded highest on these scales. This is in accordance with DBQ-studies reporting that men are involved in more violations and fewer errors than women (e.g., Reason et al., 1990). Put differently, this study showed that it is possible to make a distinction between violations and errors without letting participants complete questionnaires; however, previous factor analysis results of questionnaire responses were needed as a basis.

4.2. Speed in relation to violations and errors

Factor analysis of the MTCTs revealed a factor called paceability, which can be described as to what extent the participant's inclination for speed translates into lower MTCTs. The paceability factor distinguishes between forced-paced tasks and self-paced tasks. During forced-paced tasks, participants could not (or hardly) influence task completion times because these were computer-controlled. Participants reduced errors during these tasks, and it was found that a power curve was useful to describe the number of errors over time. During self-paced tasks, participants generally increase their speed as well as the number of violations per time unit.

Speed, errors, violations, and the number of task activations (failures + successes) were closely related. Generally, quicker participants committed more violations, made fewer errors, and were involved in more tasks per time unit. It is remarkable that quicker participants made fewer errors, considering that quicker participants performed more tasks than slower participants and that quicker driving induces higher task demands (e.g., Fuller, 2005). Being quick can therefore be seen as a sign of having good vehicle-handling abilities. Gender differences in speed-scores and error-scores were rather large when considering that gender differences in mental abilities and choice reaction time are only small (Colom et al., 2000; Lorenz & Manzey, 2001). It is likely that the average male participant had more previous practice, for example, with controlling cars in computer games or otherwise.

4.3. Results in a broader perspective

By definition, a simulator does not provide a perfect representation of real driving. Safety

has a different meaning and the consequence of committing violations is different compared to reality. Moreover, this study focused on a particular subset of people, namely youngsters around 19 years old who had the intention of obtaining their driving licence. Nonetheless, some results might be relevant to real driving behaviour on the roads. The results can clarify why violations cannot easily be prevented by practising (e.g., Hatakka et al., 2002) and why driver training cannot be considered an effective countermeasure to (speed-related) crashes (e.g., Mayhew & Simpson, 2002).

In a broader perspective, the results seem to relate to Fitts' law of human movement (Fitts, 1954). Although developed for much simpler tasks than car driving, Fitts' law states that the time needed to complete a movement depends on the level of accuracy and the human information-processing capabilities, or put differently, humans can use their capabilities to increase speed at the cost of accuracy or vice versa. In another study, such a speed-accuracy trade-off was identified with participants who practised lane-tracking on a driving simulator (De Winter et al., 2006d). The present study addressed group averages rather than the speed-accuracy trade-offs of individuals. When assuming that training on the simulator increases information-processing capabilities regarding the driving task, Fitts' law then predicts that, on average, training results in fewer errors (Figure 7) and lower task completion times (Figure 8). In a way, committing a violation is paradoxical, because violations mostly relate to the exceeding of speed-related thresholds. When considering that being quick is a sign of being competent, it requires self-control not to exceed the artificial boundary.

4.4. Implications for the development of virtual instructors

The present results can help in defining whether a failure reason is generally the result of intentional or unintentional behaviour. This knowledge can help to improve the effectiveness of virtual instructors, as it allows for task-specific feedback aimed at preventing violations or errors. A suggestion for rectifying violations in the simulator involves creating hazard awareness and establishing knowledge of traffic rules, for example, by means of small multimedia "campaigns". Errors might be prevented by putting emphasis on repeating the task and giving direct feedback on performance. In view of the fact that being quick is a sign of having good vehicle-handling abilities, it might be an interesting idea to develop training software that adapts itself to quick students so that they are automatically confronted with a more difficult task.

The violation-error structure explained only 11.9% of the total variance. Communalities of several failure reasons were low, indicating that these failures had large task-specific (unique) or random variance, unrelated to the skill for preventing errors or the tendency to commit violations. From this viewpoint, it could be considered whether tasks such as spotting objects (T14) should be removed from the curriculum. The unique variance of this task is probably not of interest, because real car driving does not involve horn pressing when objects appear and the number of failures hardly related to performance on other tasks. Second, it is recommended to research whether the assessment criteria can be improved so that the number of failures tells more about the driver. For example, stalling the engine (F1, F6, F38, F39) turned out to be a relatively rare event with low communalities and low mutual correlations, suggesting that it is something that can accidentally happen

to anyone once in a while. Other criteria might be better able to distinguish whether someone is good in driving away and stopping the car or not.

Finally, driver assessment involves more than analysing whether one follows the norms. Because (simulation-based) driving is partially a self-paced task, the assessment software should not solely concentrate on the state of the vehicle with regard to predefined thresholds, but also on the driver's pace and the number of tasks. Instead of reporting the number of failures of a participant, the number of successes should be reported too. For the same reason, it is recommended to train not only in short-lasting scenarios, which is common practice in many training simulators. Instead, training should take place in longer-lasting sessions so that students can choose speed according to their own discretion.

Appendix

This appendix provides additional information regarding the training software described in section 2.2.

Typically, the complexity of the lessons in the training programme increased progressively. So, eventually students should be able to handle complex situations and complete them without making errors or committing violations. For example, at the beginning of the first intersection lesson (lesson 6), participants had to repeatedly drive straight, turn left, and turn right, at unsignalised intersections without other traffic. Later intersection lessons (lessons 7 through 10) featured more complex tasks, including different types of signalised and sign-controlled intersections, roundabouts, and autonomous traffic.

In many of the lessons, there were other cars (agents) driving around autonomously; pedestrians and cyclists were present but did not move. Critical conflicts could occur with the agents, but there were no pre-programmed or triggered critical conflict scenarios or hazard perception scenarios such as in the work of others (e.g., Allen et al., 2007). Rather, the lessons were simulations of real-world lessons with a human instructor. In most lessons, the routes through the virtual environment were not fixed; for example, when the participant turned into a different direction than instructed by the virtual instructor, the lesson continued as normal without restarting.

The recorded tasks clearly do not represent all of the student's actions during the 15-lesson training programme; this would be unfeasible because of inherent limitations of data storage facilities (see chapters 7–10 for experiments where all data were recorded on a frequency of 50 Hz, but with a smaller number of participants and for a few tasks only). The data in this chapter represent basic performance records (success or failure and task completion time) of primary tasks that were active during the lessons, aggregated across similar lessons and/or similar types of tasks. Chapter 4 includes a factor analysis that is similar to the factor analysis in chapter 3, but in chapter 4, all recorded data were used for each primary task and each lesson block separately.

It should be noted that it is likely that the final results (i.e., the factor scores) do not depend intimately on the functional capabilities of the Green Dino Simulator, because they represent a weighted average of normalised performance records of many diverse tasks. Whereas the score on each item separately depends strongly on its definition and type of assessment, the essence of the factor scores is that they aim to be generic. It is expected that the speed-score, for example, is a personal characteristic, and that it would correlate sub-

stantially with a speed-score extracted from a different type of simulator.

More information about the simulator and the virtual instructor can be found in Fikkert et al. (2006), Green Dino (2007), Kappé and Van Emmerik (2005), Weevers et al. (2003a, 2003b), as well as in chapters 4 and 5.

5. References

- Åberg, L., & Rimmö, P.-A. (1998). Dimensions of aberrant driver behaviour. *Ergonomics*, *41*, 39–56.
- Allen, R.W., Park, G.D., Cook, M.L., & Fiorentino, D. (2007). The effect of driving simulator fidelity on training effectiveness. *Proceedings of the Driving Simulation Conference North America*, Iowa City, IA.
- Blockey, P.N., & Hartley, L.R. (1995). Aberrant driving behaviour: Errors and violations. *Ergonomics*, *38*, 1759–1771.
- Boyle, G.J. (1991). Does item homogeneity indicate internal consistency or item redundancy in psychometric scales? *Personality and Individual Differences*, *12*, 291–294.
- Colom, R., Juan-Espinosa, M., Abad, F., Garca, L.F. (2000). Negligible sex differences in general intelligence. *Intelligence*, *28*, 57–68.
- Coluccia, E., & Louse, G. (2004). Gender differences in spatial orientation: A review. *Journal of Environmental Psychology*, *24*, 329–340.
- De Winter, J.C.F., Wieringa, P.A., Kuipers, J., Mulder, J.A., & Mulder, M. (2007). Violations and errors during simulation-based driver training. *Ergonomics*, *50*, 138–158. (chapter 3 of this thesis)
- De Winter, J.C.F., Wieringa, P.A., Kuipers, J., Mulder, M., Van Paassen, M.M., & Dankelman, J. (2006). Driver assessment during self and forced-paced simulation-based training. *Proceedings of Driving Simulation Conference Europe*, Paris, France, 157–166.
- Fikkert, W., Heylen, D., Van Dijk, B., Nijholt, A., Kuipers, J., & Brugman, A. (2006). Estimating the gaze point of a student in a driving simulator. *Proceedings of the 6th International Conference on Advanced Learning Technologies*, 497–501.
- Fitts, P.M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, *47*, 381–391.
- Fuller, R. (2005). Towards a general theory of driver behaviour. *Accident Analysis & Prevention*, *37*, 461–472.
- Gorsuch, R.L. (1974). *Factor analysis*. Philadelphia: Saunders.
- Green Dino Virtual Realities (2007). Dutch Driving Simulator. <http://www.dutchsimulators.com> (Accessed 10 July 2008).
- Hatakka, M., Keskinen, E., Gregersen, N.P., Glad, A., & Hernetkoski, K. (2002). From control of the vehicle to personal self-control; broadening the perspectives to driver education. *Transportation Research Part F: Traffic Psychology and Behaviour*, *5*, 201–215.
- Kappé, B., & Van Emmerik, M.L. (2005). *The use of driving simulators for initial driver training and testing* (Report TNO-DV3 2005 C114). Soesterberg: the Netherlands, TNO Defence, Security and Safety.
- Katila, A., Keskinen, E., & Hatakka, M. (1996). Conflicting goals of skid training. *Accident Analysis & Prevention*, *28*, 785–789.

- Kontogiannis, T., Kossiavelou, Z., & Marmaras, N. (2002). Self-reports of aberrant behaviour on the roads: Errors and violations in a sample of Greek drivers. *Accident Analysis & Prevention, 34*, 381–399.
- Lajunen, T., Parker, D., & Summala, H. (2004). The Manchester Driver Behaviour Questionnaire: A cross-cultural study. *Accident Analysis & Prevention, 36*, 231–238.
- Levine, M.S. (1977). *Canonical analysis and factor comparison*. Thousand Oaks, CA: Sage Publications.
- Lin, C-T., Wu, R-C., Jung, T-P., Liang, S-F., & Huang, T-Y. (2005). Estimating driving performance based on EEG spectrum analysis. *EURASIP Journal on Applied Signal Processing, 19*, 3165–3174.
- Lorenz, B., & Manzey, D. (2001). Geschlechtsunterschiede bei Wahlreaktionsleistungen im eignungsdiagnostischen Kontext [Gender differences in choice reaction time performance in the context of aptitude testing]. *Zeitschrift für Differentielle und Diagnostische Psychologie, 22*, 173–184.
- Mayhew, D.R., & Simpson, H.M. (2002). The safety value of driver education and training. *Injury Prevention, 8*, ii3–ii8.
- Mesken, J., Lajunen, T., & Summala, H. (2002). Interpersonal violations, speeding violations and their relation to accident involvement in Finland. *Ergonomics, 45*, 469–483.
- Michon, J.A. (Ed.). (1993). *Generic Intelligent Driver Support: A comprehensive report on GIDS*. London: Taylor & Francis.
- O'Connor, B.P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, & Computers, 32*, 396–402.
- Özkan, T., & Lajunen, T. (2005). A new addition to DBQ: Positive Driver Behaviours Scale. *Transportation Research Part F: Traffic Psychology and Behaviour, 8*, 355–368.
- Parker, D., Lajunen, T., & Stradling, S. (1998). Attitudinal predictors of interpersonally aggressive violations on the road. *Transportation Research Part F: Traffic Psychology and Behaviour, 1*, 11–24.
- Parker, D., Reason, J.T., Manstead, A.S.R., & Stradling, S.G. (1995). Driving errors, driving violations and accident involvement. *Ergonomics, 38*, 1036–1048.
- Pirenne, D., Arno, P., Baten, G., & Breker, S. (2002). *TRAINER: Assessment criteria and methodology* (Deliverable 5.1). Brussels, Belgium: European Commission.
- Reason, J.T., Manstead, A., Stradling, S., Baxter, J., & Campbell, K. (1990). Errors and violations on the roads: A real distinction? *Ergonomics, 33*, 1315–1332.
- Rimmö, P.-A., & Åberg, L. (1999). On the distinction between violations and errors: Sensation seeking associations. *Transportation Research Part F: Traffic Psychology and Behaviour, 2*, 151–166.
- Rothengatter, T. (1991). Automatic policing and information systems for increasing traffic law compliance. *Journal of Applied Behavior Analysis, 24*, 85–87.
- Rothengatter, T. (1997). Errors and violations as factors in accident causation. In T. Rothengatter & E. Carbonell Vaya (Eds.), *Traffic and Transport Psychology, Theory and Application* (pp. 59–64). Amsterdam: Pergamon.
- Tabachnick, B.G., & Fidell, L.S. (2001). *Using multivariate statistics*. Boston, MA: Allyn & Bacon.
- Van Winsum, W., & Godthelp, H. (1996). Speed choice and steering behaviour in curve

- driving. *Human Factors*, 38, 434–441.
- Visser, J.A.M.M. (2001). *Het Nieuw Rijden. Handleiding voor de rijopleiding B* (Report No. 4HNR-01.10). Novem.
- Vlakveld, W.P. (2005). The use of simulators in basic driver training. *HUMANIST Workshop on the application of new technologies to driver training*, Brno, Czech Republic. Retrieved 10 July 2008 from http://www.esafetysupport.org/download/research_and_development/HUMANISTA_13Use.pdf
- Weevers, I., Kuipers, J., Brugman, A.O., Zwiers, J., Van Dijk, E.M.A.G., & Nijholt, A. (2003a). The virtual driving instructor: Creating awareness in a multiagent system. In Y. Xiang & B. Chaib-draa (Eds.), *Advances in Artificial Intelligence: 16th Conference of the Canadian Society for Computational Studies of Intelligence, Halifax, Canada* (pp. 596–602). Berlin Heidelberg: Springer-Verlag.
- Weevers, I., Nijholt, A., Kuipers, J., Van Dijk, E.M.A.G., Brugman, A.O., & Zwiers, J. (2003b). A student-adaptive system for driving simulation. *Proceedings of the 8th Australian and New Zealand Intelligent Information Systems Conference*, Sydney, Australia, 253–257.
- Xie, C., & Parker, D. (2002). A social psychological approach to driving violations in two Chinese cities. *Transportation Research Part F: Traffic Psychology and Behaviour*, 5, 293–308.