Transportation Research Part F 37 (2016) 78–96

Contents lists available at ScienceDirect

Transportation Research Part F

journal homepage: www.elsevier.com/locate/trf

The behavioral validity of dual-task driving performance in fixed and moving base driving simulators

Malte Klüver [a,c,*], Carolin Herrigel [b], Christian Heinrich [a], Hans-Peter Schöner [a], Heiko Hecht [c]

[a] Daimler AG, Driving Simulators Dept. RD/FFS, Sindelfingen, Germany
[b] MB Tech Group, Sindelfingen, Germany
[c] Department of Experimental Psychology, Johannes Gutenberg-University, Mainz, Germany

ARTICLE INFO

Article history:
Received 29 June 2015
Received in revised form 17 November 2015
Accepted 14 December 2015

Keywords:
Driving simulator
Validity
Simulator sickness
Moving-base
Fixed-base

ABSTRACT

Next generation automotive hardware and user interfaces are increasingly pre-tested in driving simulators. What are the potential limitations of such simulations? We determined the relative and absolute validity of five different driving simulators at the Daimler AG by evaluating five functions of an in-vehicle system based on the guideline of the Alliance of Automobile Manufacturers (2006). The simulations were compared to on-road driving. We hypothesized that not only simulator characteristics, but also user characteristics, such as simulator sickness, gender, or age, influence behavioral validity. Even though relating simulator characteristics and user characteristics to driving performance across different driving simulators and driving tasks is difficult, our results are surprisingly in line with the current body of research. We demonstrate the usefulness of all simulators on a relative and partially on an absolute level with moving-base simulators being preferable to fixed-base simulators. As hypothesized, we showed that simulator sickness was significantly associated with impaired performance. In the fixed-base simulators, we found a significant interaction between age and gender, which we could not find in moving-base simulators and in the on-road study. Explanations for our findings and practical implications are discussed.

1. Introduction

Conducting research using driving simulators provides several advantages, compared to real-road studies. Among others, it allows the researcher to (a) be in full control of the driving scenario, (b) examine potentially dangerous situations, (c) repeat an identical experimental condition as often as desired, (d) have access to a vast variety of data, or (e) benefit from large economic savings (Blana, 1996; Classen, Bewernitz, & Shechtman, 2011; Miller & Goodson, 1960).

However, these advantages are only of significance if the results obtained with a driving simulator can be generalized to real-world driving. Even though most validation studies to date demonstrate the usefulness of driving simulators in a variety of research questions (for review see Blana, 1996; Mullen, Charlton, Devlin, & Bédard, 2011), policy-makers still prefer real-road studies (Ranney, 2011).

In this study, for the first time, we directly compare performance in two moving-base simulator setups and three different fixed-base simulators to real-world performance in which participants accomplish secondary tasks while driving. The effect

* Corresponding author at: Daimler AG, HPC X820, 71059 Sindelfingen, Germany. Tel.: +49 151 569 751 87.
E-mail address: malte@kluver.de (M. Klüver).

http://dx.doi.org/10.1016/j.trf.2015.12.005
1369-8478/

of secondary tasks on driving performance or driver distraction received much attention due to its paramount significance for traffic safety (Young & Regan, 2007). For instance, the distracting effects of cell-phone-use (e.g. Caird, Willness, Steel, & Scialfa, 2008; Horrey & Wickens, 2006), operating a radio or cd-player (e.g. Stutts, Reinfurt, Staplin, & Rodgman, 2001), or operating a navigation system (e.g. Chiang, Brooks, & Weir, 2001) on driving performance are well known.

Several attempts have been made to establish standardized evaluation methods of in-vehicle information system (IVIS) safety (for overview see Hurts, Angell, & Perez, 2011), and recently, methodological guidelines have been developed by the Alliance of Automobile Manufacturers (AAM) (2006) and the National Highway Traffic Safety Administration (NHTSA) (2012).

One of the key purposes of this study is to determine the validity of the examined driving simulators. Mostly based on the guidelines of the AAM, we evaluated five functions of an IVIS in an instrumented vehicle as well as in five different driving simulators.

1.1. Types of driving simulator validity

In assessing driving simulator validity, two important distinctions have been established in the literature. First simulator validation approaches distinguished between physical correspondence and behavioral correspondence (Blaauw, 1982), which are nowadays mostly synonymously termed physical and behavioral validity (Mullen et al., 2011). Physical validity refers to the degree to which a simulator reproduces the physical reality. It describes to what extent the physical components such as layout, dynamic characteristics, or visual displays correspond to on-road vehicles. Thus, physical validity focuses solely on simulator characteristics. Physical validity is assumed to be higher in moving-base simulators, compared to fixed-base simulators (Godley, Triggs, & Fildes, 2002; Mullen et al., 2011). Behavioral validity, on the other hand, describes the correspondence between driving behavior in the simulator and on the real road. Thus, behavioral validity refers to the extent to which drivers behave or perform in the same manner as they do on a real road. Most researchers are in agreement about behavioral validity being the more crucial and important type of validity, compared to physical validity (e.g. Blaauw, 1982; Gemou, 2013).

Higher physical validity does not necessarily imply higher behavioral validity. For instance, for lateral motion cueing it was shown that not a scale factor of 100%, but a scale factor between 60% and 70% was preferred when regarding driving performance measures (Pretto, Nusseck, Teufel, & Bülthoff, 2009) or perceived realism (Feenstra, Wentink, Correia Grácio, & Bles, 2009). Additionally, higher fidelity of visual displays has failed to improve driving performance, compared to a lower visual fidelity (Reed & Green, 1999). In other words, there are situations in which a higher physical validity may not improve behavioral validity.

The second widely used distinction in driving simulator validity is between absolute and relative (behavioral) validity (see Blaauw, 1982). Absolute validity is established when dependent variables such as driving parameters, psychophysiological measures, or subjective evaluations take on the same numerical values in a driving simulator as in a real study. Relative validity was originally defined to be established when differences in the dependent variable between conditions are of the same order and direction (Blaauw, 1982). Some authors added to the definition that the magnitude of differences has to be identical as well (Godley et al., 2002; Mullen et al., 2011; Yan, Abdel-Aty, Radwan, Wang, & Chilakapati, 2008). Furthermore, some authors regard relative validity as a qualitative criterion (Blana, 1996), whereas others consider relative validity to be rather a quantitative criterion (Mayhew et al., 2011; Mullen et al., 2011; Wang et al., 2010). However, an explicit and unified definition of relative validity is crucially important for its statistical assessment. Only recently, Wang et al. (2010) addressed this inconsistency by clarifying that when relative validity is defined as having the same order and direction of driver differences, one refers to an identical rank ordering of conditions. However, when relative validity additionally requires the same magnitude of effects to be established, one requires differences between conditions to take on the same numerical values. Thus, it is not necessary to obtain identical numerical values to establish relative validity, but it is crucial for the intervals between conditions to be equal. In this study, we follow the latter definition by regarding relative validity to be established when the differences between conditions are numerically identical.

Whether absolute or relative validity is required depends on the purpose of a study. Since most driving studies follow an experimental design in which the effect of conditions on specific driving parameters is examined, establishing relative validity is usually sufficient (Reed & Green, 1999; Törnros, 1998). However, absolute validity is required when determining absolute numerical values, such as general take-over request times, or when intervention thresholds of an advanced driver assistance system need to be identified (Gemou, 2013). Similar validation studies regarding the safety evaluation of IVIS already exist. Wang et al. (2010) compared performance of three address entry methods in an on-road study to a fixed-base simulator, and concluded absolute validity for visual attention measures and task duration, whereas relative validity was found for lateral and longitudinal vehicle control measures.

Reed and Green (1999) examined the effect of a secondary phone task in an on-road and fixed-base simulator study on driving performance. From their findings, relative validity can be concluded for lane keeping and absolute validity can be concluded for speed variations.

Engström, Johansson, and Östlund (2005) compared performance in a secondary surrogate IVIS task in an on-road study as well as in a fixed-base and a moving-base simulator. They found that in the fixed-base simulator, lane keeping variability was substantially higher, compared to the moving-base simulator and driving on the road, even though results across environments were generally consistent. However, they found that the physiological workload and steering effort were higher in the field, compared to both simulators.

Caird et al. (2008) found in a meta-analysis that the effect of cell phone use on driving performance was not significantly different between laboratory simulators, driving simulators, and on-road studies. Thus, all simulator environments provided at least relative behavioral validity for this task.

1.2. Simulator characteristics

Most high-end driving simulators are unique and custom-built prototypes combining several individual components. Therefore, a vast and overwhelming number of parameters can be configured in such a simulator. The behavioral validity of a simulator depends on this configuration in relation to the specific driving task to be performed (Kaptein, Theeuwes, & Van Der Horst, 1996), which makes comparing the effects of simulator characteristics on driving performance across different driving simulators difficult. However, several reviews relating simulator characteristics to driving performance exist (e.g. Fischer, Eriksson, & Oeltze, 2012; Greenberg & Blommer, 2011). We chose to compare five different simulators, which primarily differ in their horizontal field of view (hFoV), their motion system, and of the type of vehicle mock-up to evaluate the impact of these factors on behavioral validity. We provide a brief selective overview of previous studies which examined these simulator characteristics and review some user characteristics before reporting our experiment.

1.2.1. Horizontal field of view (hFoV)

Generally speaking, a larger hFoV was found to be positively correlated with driving performance. Jamson (2001) compared the effect of an hFoV of 50°, 120°, and 230° on speed choice and lateral position and concluded that a higher hFoV improved the behavioral validity of the examined simulator. Kappé, van Erp, and Korteling (1999) could show that an hFoV of 150° significantly improved lane keeping performance, compared to an hFoV of 50°. Rosey and Auberlet (2014) compared driving performance between a desktop simulator with an hFOV of 93° and a fixed-base simulator including a full-scale mock-up and an hFOV of 150°. They found that the standard deviation of lane position (SDLP) and driven speed was lower, but speed variability was higher in the fixed-base simulator, compared to the desktop simulator. Jamson (2000) concluded that an hFoV of 120° does not produce more realistic speed choice behavior, but more realistic lane keeping behavior, compared to a smaller hFoV of 50°. However, Grabe, Pretto, Giordano, and Bülthoff (2010) found that a larger hFoV of 90° or 118° did not improve driving performance in a slalom-task, compared to a smaller hFoV of 45°. Pretto, Ogier, Bülthoff, and Bresciani (2009) concluded that an hFoV of 40–60° is required for valid speed estimation and that hFoVs larger than 60° do not improve speed estimation.

1.2.2. Motion system

Taken together, motion cueing in driving simulators was found in most studies to improve driving performance. More specifically, it was found that compared to no-motion conditions, motion cueing was associated with a reduced heading error and reduced number of lane violations (Greenberg, Artz, & Cathey, 2003), more realistic speed choice strategy (Reymond, Kemeny, Droulez, & Berthoz, 2001), improved performance in a slalom-driving task (Berthoz et al., 2013, but see Grabe et al., 2010), decrease of reaction times (Wierwille, Casali, & Repa, 1983), more realistic braking behavior (Siegler, Reymond, Kemeny, & Berthoz, 2001) and turning maneuvers (Hogema, Wentink, & Bertollini, 2012), and reduced driver control activity and path keeping deviations (Repa, Leucht, & Wierwille, 1982). However, it has to be emphasized that motion systems differ widely in their system capabilities and parameterizations, which makes comparisons across studies difficult.

1.2.3. Use of a mock-up

Burnett, Irune, and Mowforth (2007) compared the driving behavior in a fixed-base simulator between an "in-car"-setup utilizing a full-scale mock-up and an "out-of-car" set-up in which the participants were only provided with a full-scale dashboard. The visual display system was identical in both simulators. The participants were instructed to drive while performing several secondary tasks. Speed variability and time headway was significantly lower in the out-of-car-setup and the number of glances into the rear mirrors and the task-related touchscreen was significantly higher in the in-car setup. Furthermore, oculomotor discomfort was lower and presence ratings were higher in this case.

Rosey and Auberlet (2014) found that the speed and SDLP were higher, but the speed variability was lower in a desktop-simulator, compared to a fixed-base simulator with a full-scale mock-up. However, differences in the horizontal field of view (93° vs. 150°) make it hard to attribute the effect to the mock-up.

1.3. User characteristics

In this study, we hypothesize that behavioral validity depends not only on the simulator characteristics, but also on user characteristics. This approach becomes especially clear when considering simulator sickness. When conducting driving simulator studies, invariably some participants report symptoms such as nausea, disorientation, or oculomotor discomfort

Table 1
Participant characteristics across environments. Frequencies are given for the number of participants and gender distribution, and means are given with standard deviations in parentheses for the other variables.

	Real	MBS1	MBS2	FBS1	FBS2	FBS3
Participants	53	52	50	50	52	53
Gender (females/males)	14/39	16/36	14/36	15/35	16/36	20/33
Age	50.35 (5.33)	49.71 (5.38)	50.36 (5.56)	51.32 (4.34)	50.35 (5.33)	51.24 (4.32)
Simulator sickness	6.00 (8.31)	13.13 (17.60)	20.53 (23.84)	19.39 (22.85)	22.22 (23.29)	24.01 (35.30)
Concentration performance	145.19 (35.61)	145.27 (26.77)	142.50 (28.38)	134.96 (26.99)	145.15 (27.13)	136.36 (21.46)
Sleepiness (%)	17.83 (9.17)	28.06 (15.04)	27.85 (17.22)	27.82 (19.97)	29.41 (18.10)	26.71 (20.15)

(Kennedy, Lane, Berbaum, & Lilienthal, 1993). The usual practice with participants reporting such symptoms is including them in the analysis as long as they complete the experiment. When aborting the experiment, participants are usually dropped from the analysis. However, we doubt that this practice is appropriate, as even moderate symptoms of simulator sickness might interfere driving performance. Interestingly, Blana (1996) hypothesized that the advantages of moving-base simulators might be entirely due to reduced simulator sickness symptoms, compared to fixed-base simulators. Although there is a large body of research examining factors that contribute to the occurrence of simulator sickness (for review see Classen et al., 2011; Johnson, 2007; Kennedy & Frank, 1985; Keshavarz, Hecht, & Lawson, 2014, chap. 26; Kolasinski, 1995), it remains unclear how simulator sickness is related to driving performance. Merely a few studies have attempted to look at this relationship in flight or helicopter simulators:

Silverman and Slaughter (1995) could show in a helicopter simulator that 67% of the pilots reporting simulator sickness adjusted their flight strategy in order to minimize unpleasant maneuvers. In another helicopter study, Uliano, Lambert, Kennedy, and Sheppard (1986) found that in an air-taxi-task, simulator sickness was associated with impaired flight performance, whereas in a slalom-task, simulator sickness was not related to performance. Warner (1993) examined 18 performance parameters in two flight simulators and found that simulator sickness was significantly associated with a few, albeit not all, performance measures. Kennedy and Frank (1985) reviewed several studies examining the relationship between motion sickness and several perceptual, psychomotor, and cognitive tasks, which showed no or only a marginal negative relation between motion sickness and performance. After reviewing the literature, Money (1970) summarized that motion sickness leads to (a) decreased spontaneity, inactivity, or being quiet or subdued, (b) carelessness in performance, (c) decreased muscular coordination, (d) decreased performance in several psychomotor tasks, (e) decreased ability to estimate time and (f) decreased performance in mental arithmetic. In contrast to the deteriorating effects of simulator sickness found so far, Mullen, Weaver, Riendeau, Morrison, and Bédard (2010) could show in a driving simulator that senior drivers who did not complete an experiment due to simulator sickness showed even better driving performance, compared to drivers who completed the driving task.

Besides simulator sickness, we focus our analysis on relating age and gender to behavioral validity. To our knowledge, only Reed and Green (1999) examined gender and age differences in behavioral validity so far. They found in a dual-task paradigm that the behavioral validity of lane keeping measures depends both on age and gender. In the fixed-base simulator, participants older than 60 years produced a significantly larger SDLP, compared to the 20- to 30-year-old participants, whereas in the real study no gender or age effects were observed. Additionally, they found a significant interaction effect between age and gender in the simulator, in which older females performed significantly poorer, compared to the other groups. Regarding speed variability, they found that in the simulator the performance difference between the younger and older participants was significantly larger, compared to the real study.

1.4. The goal of this study

The study was designed to (a) determine both the relative and absolute behavioral validity in a representative use-case of five driving simulators which differ from each other in their fidelity, (b) to relate simulator characteristics to behavioral validity, and (c) to examine the effect of user characteristics on behavioral validity.

2. Methods

2.1. Participants

In this study, 215 males and 95 females (mean age 50.54, SD: 4.95) participated. They had been invited to take part in this study when holding a full driving license and self-reported to have experience with the infotainment system of Mercedes Benz (COMAND). Most participants were employees of the Daimler AG and had no prior experience with driving simulators. All participants were naive to the purpose of this study. Participants gave their informed consent and received payment for their participation. Table 1 provides an overview of the sample compositions across environments.

2.2. Apparatus and stimuli

2.2.1. Environments

The study was conducted in six different driving environments, which are illustrated in Fig. 1. In all environments but the fixed-base simulator 3 (FBS3), the current Mercedes Benz S-Class (W222) served as a vehicle or full-scale mock-up and was the basis for the vehicle dynamics model, which had been validated with the real vehicle dynamics in independent former studies. Due to technical reasons, in the FBS3 we used the vehicle dynamics model of the preceding model of the S-Class (W221). In all environments, all advanced driver assistance systems, such as adaptive cruise control or lane departure warning systems, were deactivated in the S-Class. An overview of the driving simulator characteristics is provided in Table 2.

Real vehicle (Real). In the real study, we used a Mercedes Benz S-Class in its production state. Additionally, a measurement system was installed in order to monitor the lane position and the headway, which were registered by the autonomous cruise control module. Furthermore, we attached two additional cameras to the front fenders in order to have an additional data source for the lane position. For safety reasons, a professional driving instructor accompanied the driver on the front passenger seat. The experimenter sat on the rear right seat.

Moving-base simulator (MBS1 and MBS2). The moving-base simulator at the Daimler AG is a high-fidelity driving simulator based on a 12.5-m-long linear rail system. A hexapod is mounted on top of a huge sled, which moves along this linear rail via air bearings. The hexapod comprising six linear actuators caries a spherical dome with a height of 4.5 m and an inner diameter of 7.5 m, movable in all six degrees of freedom. In the dome, a full-scale mock-up can be either positioned transversely to the linear rail (MBS1) or parallelly to the linear rail (MBS2). Hence, depending on the orientation of the mock-up, the linear axis can be fully used to simulate either lateral (MBS1) or longitudinal (MBS2) motion cues, with lateral and longitudinal scale factors of 0.7 and 0.18 for the MBS1 and 0.3 and 0.2 for the MBS2. The dome is equipped with a projection system to create a 360° horizontal field of view, with each of eight projectors having a resolution of 2048×1536 pixel, and a high-gain canvas. The side mirrors of the mock-up consist of two additional LCD displays with a resolution of 800×600 pixel on which the mirror images are generated. A more detailed specification of the moving-base simulator can be found in Zeeb (2010).

Fixed-base simulator 1 (FBS1). The fixed-base simulator 1 (FBS1) at the driving simulator center of the Daimler AG consists of a four-channel out-the-window projection system. Three channels are used to create a projection on a cylindrical 220° front screen with a radius of 3 m and a height of approximately 2.20 m. The fourth channel is used to project the rear view on a planar screen (approx. 2×2 m). Each projector's resolution is 1920×1200 pixel. The same vehicle mock-up as in the moving-base simulator was used. Hence, two additional LCD-displays were used to depict the mirror images on the side mirrors.

Fixed-base simulator 2 (FBS2). The fixed-base simulator 2 (FBS2) consists of the same mock-up and has a one channel out-the-window projection system, which projects the images with a resolution of 1920×1200 pixel and an hFoV of 82° onto a 4×3 m large screen. However, only the rear mirrors of the mock-up allow the driver a view to the rear, since the FBS2 holds no rear view projection screen.

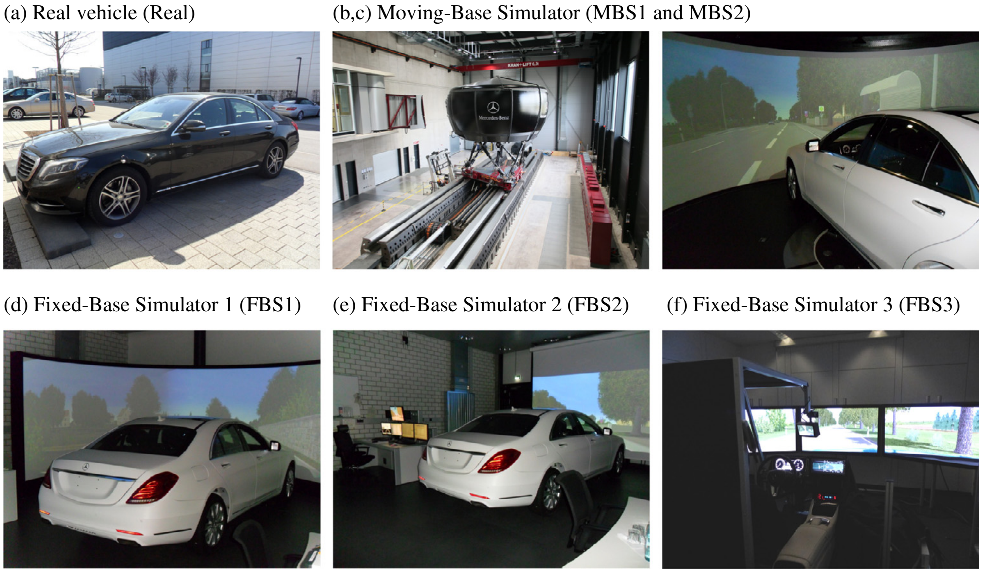

Fig. 1. Illustration of the six driving environments. Note that the MBS1 and MBS2 only differ with respect to the orientation of the full-scale vehicle mock-up.

Table 2
Overview of the driving simulator characteristics.

Simulator	Vehicle dynamics	Mock-up	Motion system	hFoV
MBS1	MB W222	Full-scale	6 DOF (lateral)	360°
MBS2	MB W222	Full-scale	6 DOF (longitudinal)	360°
FBS1	MB W222	Full-scale	None	220°
FBS2	MB W222	Full-scale	None	82°
FBS3	MB W212	Dashboard-only	None	83°

Fixed-base simulator 3 (FBS3). The third fixed-base simulator consists of three flat screens and a seat with basic control and driving instruments. The dashboard, center console, and steering wheel correspond to the interior of the current Mercedes Benz S-Class. The 65″ sized flat screens have a resolution of 1920 × 1080 pixels each and create an hFoV of 83°. A fourth LCD-display simulated the interior mirror.

2.2.2. Task types

Five functions of the in-vehicle information systems were evaluated, each associated with four specific single tasks. With the exception of the radio task, all tasks were accomplished via the COMAND system, which had to be operated with its rotary controller. In the radio task (Radio), the participants were asked to select a given frequency on a separate radio, which was installed according to the AAM-guideline (2006). Two buttons allowed the participants to increment or decrement the frequency. In the phone task (Phone), the participants had to place a call to a given contact name, which encompassed opening the phone directory in the COMAND system, finding the person from the list, and calling the contact. In the point-of-interest (POI) task, the participants had to select different terminals of the Stuttgart airport as destination, which required navigating through a hierarchical menu. The address entry task (Address) required the participants to set a given address as destination by typing both city and street name. Due to proprietary reasons, an additional task (Prototype) is not described in any detail here.

2.2.3. Route and traffic scenario

For the real study, the autobahn section between Böblingen Hulb and Rottenburg (Germany) along the A81 in both directions was chosen, which is altogether around 50 km long and holds two lanes for each direction. The experiment took place only in dry weather and daylight conditions.

The route chosen for the simulator studies was comparable in its course and curviness (mostly straight with a few curves whose radii ranged from 800 m to 1600 m). As the task was to follow a lead car, third-party vehicles were programmed such that they did not merge into the gap between the ego-car and the leading car. In contrast to the real route, the simulated route had a constant altitude profile. The traffic density of the simulated environment was manipulated to resemble real traffic as closely as possible from a subjective perspective.

In all environments, the participants followed a black leading car (Mercedes Benz B-Class, W246), which set a constant pace of 90 km/h on the right lane of the autobahn. In the real study, this was achieved by an additional experimenter who drove the leading vehicle and used its cruise control to keep the car at a constant speed.

2.3. Measures

2.3.1. Driving parameters

In the real study, all driving parameters were registered by the autonomous cruise control module. This module recorded the spatial distance between the center of the front axis and the left as well as the right inner lane markings, which allowed us to compute the SDLP. The headway was registered as the spatial distance between the front bumper of the ego-car and the rear bumper of the leading car. This distance was measured as a direct distance, disregarding the course of the road. Additionally, we recorded the driven speed in order to compute the time headway and further the standard deviation of time headway (SD Headway). A timestamp (in ms) was recorded in order to determine the completion time. Data was recorded at a sampling rate of 50 Hz. In the simulator studies, the computation of channels and sampling rate were equivalent to the computations of the autonomous cruise control module in the real study. The measurement was started when the task was set by the experimenter and stopped when the task was completed by the participant.

2.3.2. Psychological variables

Simulator sickness was measured after the drive using the Simulator Sickness Questionnaire (SSQ, Kennedy et al., 1993) and was operationalized by using the total score. Since the real study took place only during daytime (between 08:00 and 16:00 h), but in the simulator participants were driving from around 06:30 till 21:00, we additionally measured sleepiness before the experiment by a visual analogue scale with the two poles "very sleepy" and "very awake". Furthermore, we measured concentration performance before the drive by carrying out the revised test d2 (Brickenkamp, Schmidt-Atzert, & Liepmann, 2010). For the analysis, we used the raw concentration performance value (KL) of the test d2. In order to compare learning or training effects over time during the experiment across environments, we gathered the task number as a control variable, which we hereinafter term Task Order.

2.4. Design

We used a two-factorial design with Environment being a between-subjects factor and Task Type being a within-subjects factor. Environment had six factor levels (Real, MBS1, MBS2, FBS1, FBS2, FBS3). Task Type had five factor levels (Radio, Telephone, POI, Address and Prototype). For each task type, participants completed four different single tasks, all together 20 different single tasks. To avoid order effects, the order of the single tasks was randomized. As dependent variables, we measured for each task the SDLP, SD Headway, and the Completion Time. In contrast to the AAM guideline (2006), we preferred the SDLP to the number of lane exceedances, since in 87% of all observations there was no lane exceedance at all. In our opinion, this zero inflation would have produced a significant loss of information and statistical power and the high correlation between the two parameters ($r = .66$) indicates that both parameters measure a broadly similar concept. As between-subjects covariates, we gathered age, gender, simulator sickness, sleepiness, concentration performance, and as a within-subjects covariate we gathered task order.

2.5. Procedure

2.5.1. Real study

At the beginning of the study, participants filled out miscellaneous questionnaires, which, among other measures, gathered sleepiness. Subsequently, we performed the revised test d2. After reading general instructions, participants received training on how to accomplish the tasks. When they felt comfortable completing the tasks, participants familiarized themselves with the car and drove for about ten minutes until they had turned onto the autobahn. According to the AAM guidelines (2006), participants were instructed to follow the leading black Mercedes Benz B-Class and to keep their headway as constant as possible at 50 m, which equals the distance between reflective posts on German autobahns. They were further instructed to prioritize the driving task and perform the secondary task only when they believed to be able to maintain safe driving conditions. At the beginning of the autobahn drive, participants once more practiced each of the five task types before the actual experiment started and the experimenter asked the participant to perform all 20 single tasks by signaling the participants to start performing the task. The experimenter gave notice when participants successfully completed a task. After each radio task, participants were asked to set the frequency back to the initial position, which was saved in the station memory. After each COMAND task, participants were asked to return to the initial menu which acted as a starting point for the next COMAND task. An inter-task interval of at least 30 s was set. When necessary, participants were given additional time to adjust the headway again to 50 m. When a third party vehicle merged between the ego-car and leading car, the task was aborted and repeated as soon as the vehicle had left the lane, and a headway of 50 m to the leading car was set again. The duration of the drive was around 35 min. After the autobahn drive, participants were instructed to drive back to the Daimler AG driving simulation center and park the vehicle. Participants were asked to fill out several questionnaires beginning with the SSQ immediately afterward while still being seated inside the vehicle.

2.5.2. Simulator studies

The procedure in the simulator studies was the same as in the real study. Again, participants were trained and familiarized with the environment by driving along a rural road for around five minutes before joining the autobahn. Along the route, participants automatically approached and followed a Mercedes Benz B-Class, which slowly accelerated to 90 km/h and kept this pace during the whole experiment. The experimenter either announced the tasks via an intercom system or, in case of the FBS3, from the back of the room. As in the real study, participants practiced each task and were instructed in the same manner. Again, the drive took on average around 35 min. After the driving simulator session, participants directly returned to an interviewing room where they filled out several questionnaires starting with the SSQ. Around 30–60 s passed between leaving the vehicle mockup and filling out the SSQ. However, participants were instructed to indicate their symptoms in the SSQ with respect to how they had felt at the end of the simulator session.

3. Results

We drew on multilevel regression methods to analyze the data. However, multilevel analysis is rarely performed and only recently started to become widely disseminated in psychology (Baayen, Davidson, & Bates, 2008; Hoffman & Rovine, 2007). Hence, a discussion of why we chose a multilevel model to analyze the data is needed here.

Firstly, our dataset includes missing cases due to failures of the measuring system or because not all participants were able to complete all 20 tasks in the given timeframe. Traditional analysis methods, such as analysis of variance (ANOVA), require complete cases, incomplete cases are usually removed (Hox, 2010). Listwise deletion, however, may imply significant decreases in statistical power and may bias parameter estimation due to a large loss of information (Roth, 1994). In our dataset, listwise deletion would have led to the removal of 116 out of the 297 participants. Since multilevel models do not require complete cases, missing cases can easily be handled as being one of the major advantages in comparison with standard ANOVA methods (Hox, 2010; Maas & Snijders, 2003).

Secondly, another major advantage of multilevel analysis is that it allows us not only to examine the effect of between-subjects (or Level 2) covariates (in our case e.g. Simulator Sickness or Concentration Performance) as in standard analysis of covariance (ANCOVA) methods, but also the effect of within-subjects (or Level 1) covariates (in our case Task Order).

Thirdly, observations are not completely independent, as is assumed by standard regression methods, since observations within one participant are generally more similar than observations between participants. Ignoring the clustering of data might lead to deflated standard errors and fallacious significant effects. Multilevel models, however, take into account the hierarchical structure of the data (Hox, 2010).

Thus, we fitted a two-level linear regression model for each of the three dependent variables with participants on the second level and observations on the first level. On the second level, we entered the interaction effect between Environment, Age, and Gender and all its lower-order interactions and main effects as well as the main effects of Simulator Sickness, Concentration Performance, and Sleepiness as fixed effects. On the first level, we entered Task and Task Order as fixed effects. Additionally, we entered the cross-level interactions Environment × Task and Environment × Task Order as fixed effects to the model. In the random part of the model, we entered a random intercept as well as random slopes for Task and Task Order for participants. Hence, the random part corresponds to the maximal random effects structure justified by the design as recommended by Barr, Levy, Scheepers, and Tily (2013).

Before model fitting, all metric predictor variables were standardized by subjecting them to a *z*-transform. Gender was coded with males having 1 and females having −1 and Task was deviation coded with Radio being the reference level. Since the focus of the analysis was on comparing each simulator with the real study, we set the real study as the reference level for the dummy-coded Environment factor. Models were fitted using R (R Core Team, 2013) and lme4 (Bates, Maechler, Bolker, & Walker, 2014). After model fitting, we visually inspected the residual plots. Since the residuals of the models predicting the SDLP, SD Headway and completion time were not normally distributed, as it is assumed in regression models, we took the logarithm of all three dependent variables and fitted the models again. Since outliers can lead to inflated error rates and bias parameter estimation (Osborne & Overbay, 2004), we subsequently removed observations with a standardized residual at a distance greater than 2.5 standard deviations from 0 and fitted the models again on the outlier-cleaned datasets. A repeated visual inspection of residual plots revealed no deviation from homoscedasticity or normality for the residuals and the random intercepts and slopes.

We chose to present the results in the form of an ANOVA table, since it is more widespread in psychology and interpreting regression coefficients with dummy variables and interactions is somewhat tricky (Brambor, Clark, & Golder, 2006). However, we provide the regression output in Appendix A of this Paper. We used the afex package (Singmann & Bolker, 2014) in order to compute the Type III ANOVAs based on the multilevel models and we used the Kenward-Rogers approximation for the denominator degrees of freedom. Results are given in Table 3.

To further illustrate the significant main and interaction effects, we simulated expected values based on the multilevel regression models as proposed by King, Tomz, and Wittenberg (2000). In the following section, results are presented for each dependent variable separately.

3.1. Standard deviation of lane position (SDLP)

The simulated expected values illustrated in Fig. 2 indicate that the SDLP across tasks was similar between the real study and the moving-base simulators. However, the fixed-base simulators clearly differed from the real study. Indeed, simulated first differences according to King et al. (2000) revealed that the real study and the MBS2 did not significantly differ from each other ($z = 1.792$, $p = .073$). The MBS1 elicited a significantly lower SDLP, compared to the real study, only at an uncorrected α-level ($z = 2.209$, $p = .027$), whereas in comparison to the real study, the SDLP was significantly higher in the FBS1 ($z = 5.357$, $p < .001$), FBS2 ($z = 7.552$, $p < .001$), and FBS3 (6.584, $p < .001$) at a Bonferroni-corrected α-level of 0.01.

Regarding gender and age effects, we found that performance was about equal in the real study and the moving-base simulators. In the fixed-base simulators, however, we found that the older participants performed poorly, compared to the younger participants, and that especially older females had a larger SDLP, in comparison to the other groups.

All simulators but the FBS3 significantly differed from the relationship between task order and SDLP in the real study, as can be inferred from the regression table in Appendix A. A significant performance improvement between first and last task was found only for the FBS1 ($z = 2.934$, $p = .003$) and for the FBS2 ($z = 4.450$, $p < .001$).

Concentration performance was highly associated with a smaller SDLP, whereas simulator sickness and sleepiness had no effect on the SDLP.

3.2. Standard Deviation of Headway (SD Headway)

Simulated expected values for the SD Headway are illustrated in Fig. 3. They clearly show an offset of all simulators relative to the real study. The moving-base simulators had an offset of around 0.06 s, whereas the fixed-base simulators produced about twice as large SD Headways, compared to the real study. Simulated first differences revealed that the SDLP in the real study was significantly lower, compared to the MBS1 ($z = 5.915$, $p < .001$), MBS2 ($z = 5.176$, $p < .001$), FBS1 ($z = 9.226$, $p < .001$), FBS2 ($z = 10.707$, $p < .001$), and FBS3 ($z = 8.652$, $p < .001$).

Table 3
ANOVA table from the multilevel regression analysis for all three dependent variables.

Dependent variable	Effect	df_1	df_2	F	p
log(SDLP)	Environment	5	271.68	29.77	<.001***
	Task	4	274.63	168.47	<.001***
	Gender	1	266.14	6.18	.01**
	Age	1	270.61	21.42	<.001***
	Simulator sickness	1	271.85	.24	.62
	Concentration performance	1	259.32	15.28	<.001***
	Sleepiness	1	258.93	2.29	.13
	Task order	1	268.93	13.64	<.001***
	Gender × Age	1	270.52	.75	.39
	Environment × Task	20	595.43	.86	.64
	Environment × Gender	5	266.06	.64	.67
	Environment × Age	5	267.42	2.55	.03*
	Environment × Task Order	5	268.35	10.05	<.001***
	Environment × Gender × Age	5	267.25	2.39	.04*
log(SD Headway)	Environment	5	266.46	25.67	<.001***
	Task	4	274.23	219.78	<.001***
	Gender	1	265.02	11.65	<.001***
	Age	1	268.89	37.61	<.001***
	Simulator sickness	1	267.89	5.52	.02*
	Concentration performance	1	259.34	12.65	<.001***
	Sleepiness	1	259.01	.14	.71
	Task order	1	270.12	12.08	<.001***
	Gender × Age	1	266.84	3.35	.07
	Environment × Task	20	592.16	1.45	.09
	Environment × Gender	5	265.20	1.76	.12
	Environment × Age	5	266.20	1.84	.11
	Environment × Task Order	5	269.57	1.16	.33
	Environment × Gender × Age	5	266.05	2.65	.02*
log(Completion Time)	Environment	5	272.13	6.85	<.001***
	Task	4	273.77	3603.75	<.001***
	Gender	1	266.37	31.98	<.001***
	Age	1	269.98	16.41	<.001***
	Simulator sickness	1	267.35	5.22	.02*
	Concentration performance	1	259.84	21.52	<.001***
	Sleepiness	1	258.39	.50	.48
	Task order	1	266.27	67.81	<.001***
	Gender × Age	1	269.59	.11	.74
	Environment × Task	20	591.30	3.45	<.001***
	Environment × Gender	5	266.14	.71	.62
	Environment × Age	5	266.96	.30	.91
	Environment × Task Order	5	265.78	7.49	<.001***
	Environment × Gender × Age	5	266.78	1.26	.28

* $p \leqslant .05$.
** $p \leqslant .01$.
*** $p \leqslant .001$.

As with the SDLP, performances in the real study and in the moving-base simulators were mostly independent of age and gender. Again, we found a strong interaction effect between age and gender across the fixed-base simulators, especially with older females having a significantly higher SDLP than the other groups.

Simulator sickness as well as concentration performance were significantly correlated with SD Headway. Task Order was significantly associated with SD Headway, but this effect was similar across Environments. Sleepiness had no significant effect.

3.3. Completion time

The illustrated effects in Fig. 4 show that all simulators had an offset, compared to the real study. Indeed, the task completion time was significantly higher in the MBS1 ($z = 2.733$, $p = .006$), MBS2 ($z = 3.481$, $p < .001$), FBS1 ($z = 3.586$, $p < .001$), FBS2 ($z = 5.231$, $p < .001$), and FBS3 ($z = 5.505$, $p < .001$).

Except in the FBS3, we could find a tendency of older females having a longer completion time, compared to the other groups. However, the regression table in Appendix A reveals that only in the FBS2 this interaction was significantly different from the real study.

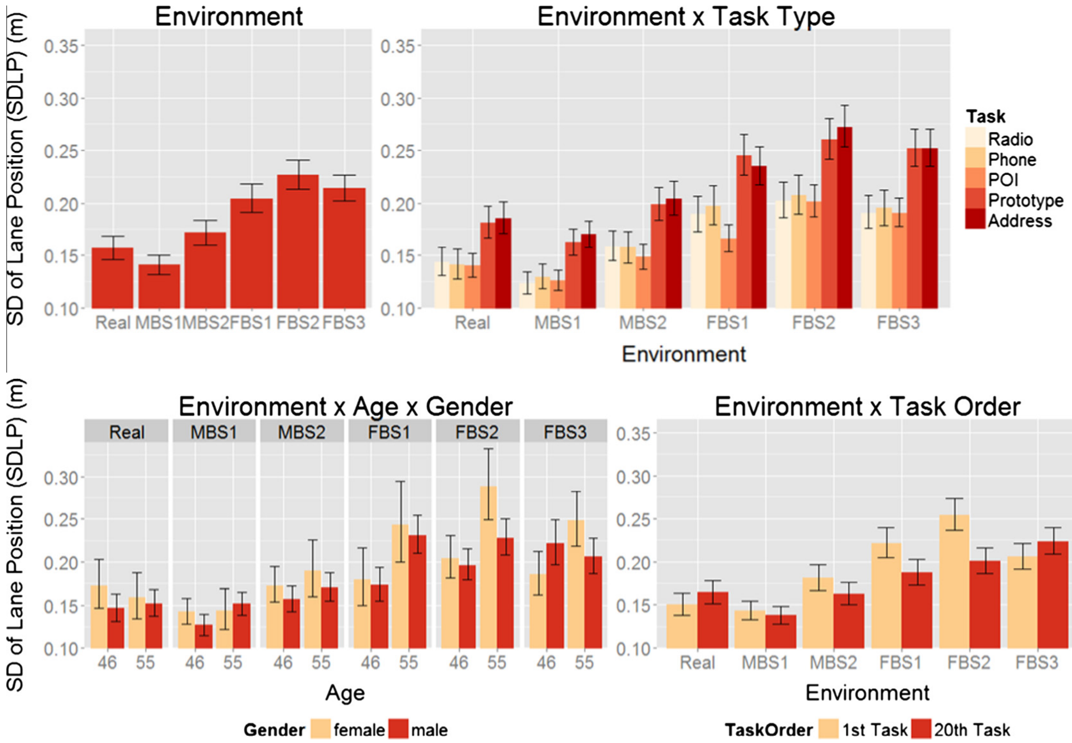

Fig. 2. Simulated expected values illustrating the main effect of Environment and the interaction effects of Environment × Task Type, Environment × Age × Gender, and Environment × Task Order on the SDLP while holding all other predictor variables at their mean. Error bars indicate 95% confidence intervals. For age, we chose predictions for 46 and 55 year-old participants, which correspond to one standard deviation below and above the mean age.

Simulator sickness was positively related to completion time. In all environments, but the FBS3, the participants improved their performance during the experiment. Compared to the real study, the performance improvement was significantly stronger in the FBS2 and the effect of Task Order in the FBS3 was significantly different from the real study as inferred from the regression table in Appendix A.

3.4. Quantifying absolute and relative validity

In line with Wang et al. (2010), we regard relative validity to be established by the absence of the interaction between Environment and Task Type. Absolute Validity is established when both the interaction effect and the main effect of Environment are absent or approach zero.

Since the regression coefficients can only be interpreted as single contrasts to the intercept (see Brambor et al., 2006), we quantified both absolute and relative validity by computing the effect sizes for both main- and interaction effects on the simulated expected values with the factors Environment (levels: Real vs. Simulator) and Task Type (levels: Radio, Phone, POI, Prototype, Address) for each simulator and each dependent variable while holding all other predictors at their mean. Note that in each case the dependent variable remained logarithmic.

To conclude relative validity, no or little variance should be explained by the interaction effect between Environment and Task. The variance explained by Environment provides information about the size of the offset, with low (high) variance meaning a small (large) offset. Absolute Validity is established, when both the main effect of Environment and the interaction effect between Environment and Task Type are absent or approach zero. Thus, most of the variance should ideally be explained by the Task Type. Results are given in Table 4.

Even though some interaction effects were statistically significant in the multilevel analysis, as presented in the regression table in Appendix A, effect sizes, as shown in Table 4 indicate, that the interaction effects are only of marginal relevance. For almost all simulators and measures, the interaction effects explained less variance, compared to the residual error term. Hence, we believe that concluding relative validity for all simulators regarding all dependent variables of this study is justified.

However, simulators differed largely in their absolute validity. For the SDLP, between 46% and 64% of the variance of the fixed-base simulators can be explained by an offset relative to the real road condition. Hence, absolute validity cannot be

Fig. 3. Simulated expected values illustrating the main effect of Environment and Simulator Sickness as well as the interaction effects of Environment × Task Type and Environment × Age × Gender on SD Headway while holding all other predictor variables at their mean. Error bars and error bands indicate the 95% confidence intervals. For age, we chose predictions for 46 and 55 year-old participants, which corresponds to one standard deviation below and above the mean age.

concluded for these simulators. In contrast, the offset for the MBS2 explains only a marginal proportion of the variance in comparison with the residuals, thus concluding absolute validity can be seen as justified. In the MBS1, the offset explains more variance, compared to the MBS2. However, it is a matter of argument, whether absolute validity can be concluded. Among fixed-base simulators, FBS1 had the smallest offset but the largest interaction effect. The FBS3 had a smaller offset, compared to FBS2.

Regarding SD Headway, absolute validity cannot be concluded for any simulator. However, the moving-base simulators produced a substantially smaller offset, which explains around 30–35%, compared to the offset of the fixed-base simulators, which explains almost twice as much variance (i.e. 50–60%). When comparing the fixed-base simulators, the FBS3 had the smallest offset, but the largest interaction effect, whereas FBS1 and FBS2 were comparable.

Around 95% of the completion time variance was accounted for by the Task Type in all simulators. Hence, the Environment as well as the interaction effect explained only a marginal proportion of the overall variance. Again, it is debatable whether 2–5% of explained variance are of practical relevance. We believe that concluding absolute validity for the MBS1, MBS2, and FBS1 can be regarded as justified.

4. Discussion

The goal of this validation study was to determine the behavioral validity of five different driving simulators and to relate both driving simulator characteristics and user characteristics to behavioral validity. In the following section, we discuss the results in the light of these aspects. Furthermore, we formulate practical implications from our work.

4.1. Absolute and relative behavioral validity

We determined the behavioral validity of five different driving simulators. The highest behavioral validity was accomplished with the two moving-base configurations MBS1 and MBS2. For both, we could conclude absolute validity for the SDLP and for the completion time. However, for the headway variability we could conclude only relative validity. Compared to the fixed-base simulators, the moving-base simulators are considerably superior in the absolute validity of the headway variance. For all fixed-base simulators, we conclude relative, but not absolute behavioral validity for all dependent measures.

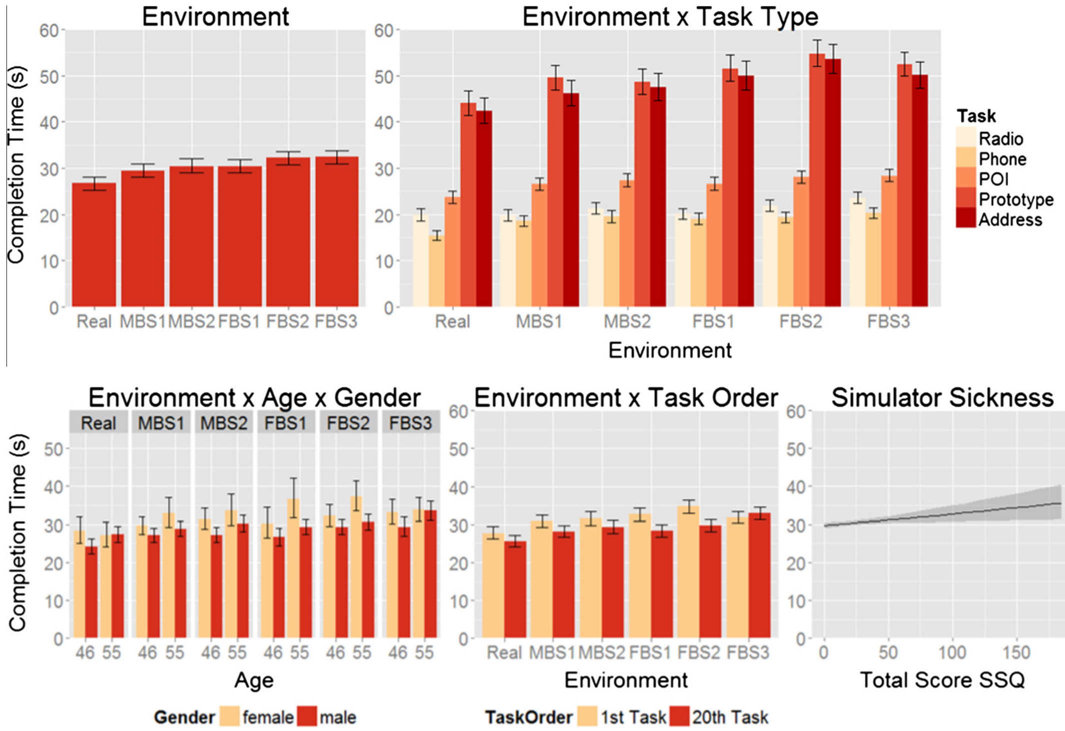

Fig. 4. Simulated expected values illustrating the main effect of Environment and Simulator Sickness, and the interaction effects of Environment × Task Type, Environment × Age × Gender, and Environment × Task Order on the completion time while holding all other predictor variables at their mean. Error bars and error bands indicate the 95% confidence interval. For age, we chose predictions for 46 and 55 year-old participants, which correspond to one standard deviation below and above the mean age.

Table 4
Effect sizes η^2 for the main and interaction effects of Task Type and Environment. Effect sizes are given as the percentage of explained variance. Note that tests for statistical significance were not performed since the computation is based on simulated data, and not on the raw data, and η^2 is independent of sample size.

Variable	Effect	MBS1	MBS2	FBS1	FBS2	FBS3
SDLP	Environment	13.32%	9.90%	46.29%	64.43%	56.21%
	Task	77.56%	80.26%	46.54%	31.95%	39.45%
	Environment × Task	0.67%	0.33%	2.03%	0.09%	0.14%
	Residuals	8.45%	9.52%	5.14%	3.53%	4.2%
Headway SD	Environment	34.60%	30.41%	57.13%	60.01%	50.69%
	Task	59.75%	62.63%	39.02%	37.83%	45.43%
	Environment × Task	2.19%	3.01%	1.47%	0.34%	1.52%
	Residuals	3.46%	3.94%	2.38%	1.83%	2.37%
Completion time	Environment	1.43%	2.62%	2.34%	4.52%	5.41%
	Task	97.45%	96.28%	96.49%	94.63%	93.84%
	Environment × Task	0.57%	0.50%	0.65%	0.38%	0.23%
	Residuals	0.54%	0.59%	0.52%	0.48%	0.52%

Table 5
Concluded validity types for the examined simulators regarding all dependent measures.

Simulator	SDLP	SD Headway	Completion time
MBS1	Absolute validity[a]	Relative validity	Absolute validity[a]
MBS2	Absolute validity	Relative validity	Absolute validity[a]
FBS1	Relative validity	Relative validity	Absolute validity[a]
FBS2	Relative validity	Relative validity	Relative validity[a]
FBS3	Relative validity	Relative validity	Relative validity[a]

[a] Relative validity can be clearly concluded, whereas it is arguable whether absolute validity can be concluded.

However, it is a moot point to discuss whether absolute validity can be concluded for the completion time. Fixed-base simulators differed only marginally from each other. An overview of the concluded validity types for the simulators is given in Table 5.

Our results are consistent with the results from previous studies, which concluded relative validity for the examined simulators regarding lateral and longitudinal measures (Caird et al., 2008; Engström et al., 2005; Reed & Green, 1999). At the same time, the results point to the importance of the moving base in order to obtain maximal absolute validity.

Taken together, all simulators proved to be useful research tools. What is more, the MBS1 and MBS2 were found partially capable to fully replace real studies even on an absolute scale.

4.2. Simulator characteristics

4.2.1. Horizontal field of view (hFoV)

The effect of the hFoV can be clearly isolated by comparing the FBS1 and the FBS2, since these two driving simulators only differed in their hFoV (220° vs. 82°). For lateral and longitudinal control as well as completion time, the FBS1 seemed to elicit a slightly better and more realistic driving performance, compared to the FBS2, which is in line with the results from Jamson (2000), Jamson (2001), Kappé et al. (1999), and partially with Rosey and Auberlet (2014). However, our findings are in contrast to Grabe et al. (2010) who found no improved slalom-task performance for hFoVs of 90° and 118°, compared to an hFoV of 45°. However, in comparison to the advantage of the moving-base, the effect of hFoV on behavioral validity is rather small.

4.2.2. Motion system

In line with previous studies, we found that a moving base significantly improves driving performance and increases behavioral validity (e.g. Berthoz et al. 2013; Greenberg et al. 2003; Repa et al. 1982). This is not surprising since the secondary tasks required the participants to shift their visual attention during the task away from the road to the IVIS. Since the visual channel in fixed-base simulators is the only relevant source for vehicle control, such a shift implies controlling the vehicle without any perceptual input, whereas in moving base simulators, participants can rely at least on the remaining vestibular cues, and likely on kinesthetic and tactile cues as well. Interestingly, we found that the MBS1 had an even more precise lanekeeping control, compared to the real study.

Not surprisingly, we could show that driving performance depends clearly on the parametrization and capabilities of the motion system. For lateral motion cueing, the MBS1 has a motion space of about 12 m. When considering the width of a two- or three-lane highway, this motion space is in general sufficient to simulate motion cues for lane change or simple lane keeping maneuvers. However, considering the distance traveled when one is accelerating from 0 to 100 km/h, one easily realizes that 12 m are not sufficient to simulate the longitudinal motion cues adequately. Therefore, the motion system had to be scaled down considerably compared to the linear accelerations found in real-world driving, so that the participants might not have been able to accurately interpreted these reduced motion cues. This would explain why the lateral movement provided by the MBS1 was superior to the MBS2, whereas headway variability in the MBS2 was only slightly higher, compared to the MBS1, but not identical with the real study.

However, another possible explanation for this finding is that in the real study, the participants might have concentrated on avoiding third party cars merging into the gap by keeping their headway much shorter, compared to those in the simulator studies, which is concomitant with a reduced headway variability. Beyond question, driving is a social task, and driving behavior considerably changes depending on whether the surrounding traffic is regarded as real or virtual (Mühlbacher, Rittger, & Maag, 2014). A further analysis of the data supports this notion: The average minimum as well as the average maximum of headway time over tasks of the participants was significantly lower in the real study, compared to the simulator studies.

Interestingly, even though the MBS1 elicits a more precise lanekeeping behavior, compared to the MBS2, the MBS2 tends to elicit a slightly more realistic lanekeeping behavior. On the one hand, one could conclude from this finding that lateral motion cues in the MBS1 are exaggerated. On the other hand, it can be argued that the traffic scenario in the simulators was more predictable and a larger SDLP in the real study might have been due to a more complex traffic scenario, such as merging or overtaking third-party vehicles. Nevertheless, both moving-base simulators did not differ significantly from the real study.

4.2.3. Use of a mock-up

The most significant difference between the FBS2 and FBS3 is the existence of a full-scale mock-up. The results show that the FBS3 is slightly better, regarding SDLP and headway variability, whereas completion time was comparable. These results are in line with Burnett et al. (2007), who found that a full-scale mockup produced higher variance of longitudinal control, compared to a dashboard-only simulator.

Moreover, our results are not at variance with the results from Rosey and Auberlet (2014). They compared SDLP and speed variability across a 93° dashboard-only simulator and a 150° fixed-base simulator including a full-scale mock-up. Their simulator setups would correspond, to some extent, to the FBS1 and FBS3 in our study. Consistent with their results, we found the SDLP to be slightly lower in the FBS1, compared to the FBS3, but for longitudinal control variability we found the FBS3 to be slightly better. Taken together, it is likely that a mock-up alone impairs both lateral and longitudinal control, compared to a dashboard-only setup. However, the improved lateral control in the fixed-base simulator found by Rosey and Auberlet (2014) might be explained by a compensating effect of the larger hFoV.

4.3. User characteristics

As hypothesized, we found that simulator sickness affects simulator validity. Simulator sickness was significantly related to a higher headway variability and longer completion times. However, simulator sickness was not associated with lateral control measures. Our finding of simulator sickness being associated with rather impaired performance measures are consistent with most previous studies from flight or helicopter studies (Kennedy & Frank, 1985; Money, 1970; Silverman & Slaughter, 1995; Uliano et al., 1986; Warner, 1993). However, our results are at variance with the results from Mullen et al. (2010), who found that experiencing simulator sickness was correlated with improved driving performance in senior drivers. To sum up, it is likely that simulator sickness goes along with an impairment of some, but not all, performance measures as found previously (Kennedy & Frank, 1985; Uliano et al., 1986; Warner, 1993).

However, the basis of this relationship remains unclear. Few empirical findings so far have yielded information about the causal direction of this relationship, for instance whether simulator sickness leads to performance decrease or whether a specific driving performance leads to simulator sickness. Only a few experimental studies on this relationship exist. Reviewed by Kennedy and Frank (1985), they were conducted in the 1940s and indicate that simulator sickness precedes impaired performance.

Furthermore, it is conceivable that the relationship is mediated by third variables, such as motivation, as suggested by Money (1970), or that driving performance and simulator sickness mutually reinforce each other. Further research is necessary to understand the underlying causal relationship.

We found no or only marginal age and gender differences between the real study and the moving-base simulators. Interestingly, in the fixed-base simulators, the older participants showed a considerably higher headway and lane keeping variability, whereas the performance of the younger participants was roughly the same as in the real study. Furthermore, especially the older female participants showed considerable difficulties in lateral and longitudinal control in the fixed-base simulators.

Such an effect was already found by Reed and Green (1999) and is strikingly similar to our findings. Reed and Green (1999) additionally compared performance between a real study and a simulator study in a normal driving task without secondary tasks and found that all age and gender groups performed similarly in the simulator.

Taking these findings together, the gender-age interaction is only present in a fixed-base simulator, but not in a moving-base simulator or in the real study and is only present in a dual-task paradigm, but not in a single driving task. How can this be explained?

Lesch and Hancock (2004) found that especially older females were more affected by a secondary phone-task, compared to the other age and gender groups. They explained their findings by assuming that older females in particular were less aware of their decreased performance compared to the other groups. However, in this case we would have expected that this gender - age difference is not limited to fixed-base simulators only.

Salthouse (1996) hypothesized that age-related performance decrements can be explained by a generally reduced speed of cognitive processing. However, in this case we would again have expected age effects in all environments. Additionally, since we statistically controlled for concentration performance, it is unlikely that a general slowdown of cognitive processes can explain this effect.

In a meta-analytic review, Riby, Perfect, and Stollery (2004) could show that age-differences in dual-task performances strongly depend on the task domain. They found that with age, dual-task performance only deteriorates when tasks require controlled or motor processes, whereas automatic processes stay relatively intact. These findings provide a promising explanation for our results. In the real study as well as in the moving-base simulators, the participants were provided with visual and vestibular cues to control the vehicle. Under these conditions, vehicle control can be expected to rely mostly on automatic processes since driving is likely to be a highly learned task. In the fixed-base simulators, however, the driving task is solely based on visual cues, which might require higher cognitive processes to compensate for the missing vestibular cues. Furthermore, in the real as well as in the moving-base simulator studies, the participants were able to simultaneously perform both tasks by automatically processing vestibular cues when visually focusing on the secondary task, whereas in the fixed-base simulators visual attention could only be directed exclusively either to the primary or secondary task, which might have required higher cognitive control strategies. However, this approach only explains the age but not the gender differences.

Furthermore, we believe it is worth mentioning that the age and gender differences strikingly correspond to the age- and gender-differences in the usage frequency of new technologies in Germany (Statistisches Bundesamt, 2006). Thus, it is conceivable that prior experience with computer technologies might be the basis of this effect. In order to explain why no age and gender effects were observed in moving-base simulators, it is conceivable that a moving-base simulator provides a much more realistic driving experience, so that a much smaller transfer from reality to a moving-base simulator is required, compared to a fixed-base simulator. Thus, experience with new technologies might facilitate adaptation to virtual environments. Since behavioral correspondence between the on-road study and the simulator can be regarded as a behavioral measure of the sense of presence, Lachlan and Krcmar (2011) provide empirical support for this explanation by showing that prior gaming or computer knowledge was positively correlated with the sense of presence. However, Felnhofer et al. (2014) failed to find this relationship. Furthermore, it was found that males report a higher sense of presence than females (Felnhofer, Kothgassner, Beutl, Hlavacs, & Kryspin-Exner, 2012; Nicovich, Boller, & Cornwell, 2005; Slater, McCarthy, & Maringelli, 1998, but see Felnhofer et al., 2014) and that younger adults adapt faster, compared to older adults, to virtual environments (Schellenbach, Lövdén, Verrel, Krüger, & Lindenberger, 2010), which corresponds to our findings. Nonetheless, the basis of this effect remains theoretically challenging and requires further empirical support.

4.4. Practical implications

Several practical recommendations can be drawn from our results. Firstly, we found that simulator sickness was significantly related to performance measures. Thus, since already moderate simulator sickness symptoms influence performance, we recommend to statistically control for simulator sickness in the analysis of performance measures, as already proposed by Stoner, Fisher, and Mollenhauer (2011).

Secondly, when evaluating the safety of IVIS, moving-base simulators are clearly preferable to fixed-base simulators. Not only do they elicit performance which is much closer to on-road performance, but they are also far more independent from user characteristics. Thus, moving-base simulators can reveal differences among tasks more precisely, whereas in fixed-base simulators, much of the variance would remain unexplained due to large gender and age differences. Furthermore, our results show that moving-base simulators are not superior to fixed-base simulators simply due to differences in simulator sickness, as hypothesized by Blana (1996), since we found no substantial differences in simulator sickness across driving simulators (Klüver, Herrigel, Preuß, Schöner, & Hecht, 2015).

Thirdly, since we found strong learning effects in the fixed-base simulators, there is a chance that more valid performance can be elicited when simulator training is provided to participants before the experiment. Prior training was found to improve driving performance significantly (Hill & Salzman, 2012; Roenker, Cissell, Ball, Wadley, & Edwards, 2003) and a standardized training procedure was already proposed by Hoffmann and Buld (2006). Thus, it is likely that a more elaborate training increases behavioral validity in fixed-base simulators. In moving-base simulators, however, our results do not indicate the need for a more elaborate simulator training.

Fourthly, from an economic perspective, institutions are particularly interested in the most efficient use of their driving simulators with short downtimes. Thus, at the Daimler AG we schedule experiments from the early morning hours to the late evening. Since we could not find a relation between sleepiness and performance measures, our results do not support concerns regarding the usefulness of data from very early or late experimental sessions.

5. Conclusion

We have evaluated the performance across five different functions of the IVIS in six different environments. Our results show that all driving simulators are comparable to the real study on a relative scale, thus we could demonstrate the usefulness of all simulators. Even though it is very difficult to compare performance measures across driving simulators and driving tasks, we provide the first comprehensive comparison of multiple platforms. Moving base simulation most closely approaches real-world driving. Note that this effect was found on top of the positive effect of horizontal field of view. Our results are astonishingly consistent with previous studies examining the effect of driving simulator characteristics and user characteristics on behavioral validity. Furthermore, we could show that user characteristics play a substantial role in the behavioral validity: We found that simulator sickness was significantly associated with impaired performance and that gender and age substantially influenced performance in fixed-base simulators. However, future research is necessary, to understand the underlying mechanisms of these effects.

Acknowledgments

The authors wish to thank Nadine Riemer and Olga Steinhauer for their assistance in collecting the data. We thank the Daimler AG Research and Advanced Engineering Unit for their support.

Appendix A

Results from the multilevel regression analysis for all three dependent variables. For each predictor, beta coefficients are given with standard errors in parentheses. Note that Task is deviation coded and Environment is dummy coded, so that the intercept holds the grand mean for the Real Environment with all other predictors held at their mean. All metric predictor variables were centered and *z*-standardized.

Fixed effects	Dependent variable		
	log(SDLP)	log(Headway SD)	log(Time)
(Intercept)	−1.851 (.037)***	−2.118 (.054)***	3.284 (.027)***
Task (Reference: Radio)			
Phone	−.106 (.029)***	−.384 (.050)***	−.550 (.018)***
POI	−.115 (.027)***	−.092 (.045)*	−.121 (.015)***
Prototype	.144 (.025)***	.418 (.043)***	.501 (.016)***
Address	.167 (.025)***	.322 (.043)***	.461 (.017)***
Environment (Reference: Real)			
MBS1	−.109 (.050)*	.426 (.073)***	.100 (.037)**
MBS2	.091 (.051)	.382 (.075)***	.132 (.038)***
FBS1	.264 (.050)***	.669 (.073)***	.131 (.037)***
FBS2	.367 (.049)***	.763 (.072)***	.187 (.036)***
FBS3	.310 (.047)***	.607 (.069)***	.193 (.035)***
Environment × Task			
MBS1 × Phone	.021 (.039)	.199 (.067)**	.088 (.024)***
MBS1 × POI	.003 (.036)	−.067 (.060)	.017 (.020)
MBS1 × Prototype	.000 (.033)	−.103 (.058)	.016 (.021)
MBS1 × Address	.022 (.034)	−.039 (.058)	−.014 (.022)
MBS2 × Phone	.017 (.040)	.187 (.069)**	.104 (.024)***
MBS2 × POI	−.031 (.037)	−.064 (.062)	.014 (.021)
MBS2 × Prototype	.000 (.034)	−.175 (.059)**	−.034 (.022)
MBS2 × Address	.005 (.035)	.011 (.059)	−.018 (.023)
FBS1 × Phone	.069 (.040).	.202 (.069)**	.080 (.025)***
FBS1 × POI	−.091 (.037)*	−.050 (.062)	−.015 (.021)
FBS1 × Prototype	.037 (.035)	−.092 (.060)	.026 (.022)
FBS1 × Address	−.028 (.035)	−.072 (.060)	.034 (.023)
FBS2 × Phone	.015 (.039)	.096 (.067)	.036 (.024)
FBS2 × POI	−.002 (.036)	−.058 (.060)	−.017 (.020)
FBS2 × Prototype	−.005 (.034)	−.057 (.058)	.031 (.021)
FBS2 × Address	.016 (.034)	.016 (.058)	.048 (.023)*
FBS3 × Phone	.016 (.039)	.124 (.067).	.080 (.024)***
FBS3 × POI	−.001 (.036)	−.139 (.061)*	−.011 (.020)
FBS3 × Prototype	.019 (.033)	−.051 (.058)	−.019 (.021)
FBS3 × Address	−.004 (.034)	−.054 (.058)	−.024 (.022)
Age and Gender			
Gender	−.051 (.035)	−.090 (.053).	−.039 (.026)
Age	−.012 (.035)	.032 (.052)	.019 (.025)
Gender:Age	.031 (.034)	.103 (.051)*	−.003 (.011)
Environment × Gender			
MBS1 × Gender	.034 (.047)	.074 (.071)	−.018 (.035)
MBS2 × Gender	−.002 (.048)	.067 (.073)	−.024 (.036)
FBS1 × Gender	.029 (.048)	.002 (.072)	−.048 (.035)
FBS2 × Gender	−.018 (.046)	−.099 (.070)	−.035 (.034)
FBS3 × Gender	.049 (.045)	.057 (.068)	.005 (.033)
Environment × Age			
MBS1 × Age	.060 (.043)	.072 (.065)	.023 (.032)
MBS2 × Age	.057 (.044)	.102 (.067)	.024 (.033)
FBS1 × Age	.160 (.057)**	.267 (.085)**	.052 (.041)
FBS2 × Age	.136 (.043)**	.072 (.065)	.028 (.032)
FBS3 × Age	.068 (.047)	.105 (.071)	.019 (.034)
Environment × Gender × Age			
MBS1 × Gender × Age	.010 (.043)	−.132 (.065)*	−.055 (.032)
MBS2 × Gender × Age	−.033 (.044)	−.115 (.067)	−.033 (.033)

(continued on next page)

Appendix A (*continued*)

Fixed effects	Dependent variable		
	log(SDLP)	log(Headway SD)	log(Time)
FBS1 × Gender × Age	−.034 (.056)	−.164 (.084)	−.066 (.041)
FBS2 × Gender × Age	−.078 (.043)	−.230 (.065)***	−.067 (.032)*
FBS3 × Gender × Age	−.122 (.047)**	−.209 (.070)**	−.013 (.034)
Environment × TaskOrder			
MBS1 × TaskOrder	−.039 (.017)*	.010 (.031)	.042 (.025)
MBS2 × TaskOrder	−.059 (.018)***	−.009 (.032)	.002 (.011)
FBS1 × TaskOrder	−.076 (.018)***	−.019 (.033)	−.019 (.011)
FBS2 × TaskOrder	−.097 (.018)***	−.017 (.032)	−.023 (.011)*
FBS3 × TaskOrder	−.002 (.018)	.045 (.032)	.033 (.011)**
TaskOrder	.026 (.013)*	−.034 (.023)	−.024 (.008)**
D2	−.052 (.013)***	−.072 (.019)***	−.046 (.009)***
SSQ	.007 (.013)	.047 (.019)*	.023 (.009)*
Sleepiness	−.021 (.013)	−.008 (.020)	.007 (.010)
Random effects			
Participant variances			
(Intercept)	0.035	0.067	0.02
TaskOrder	0.001	0.004	<.001
Phone	0.01	0.03	0.004
POI	0.005	0.013	0.001
Prototype	0.001	0.006	0.001
Address	0.002	0.006	0.003
Model fit			
Removed outliers	86	101	138
Number of observations	5308	5293	5256
Participants	288	288	288
Log likelihood	−2156	10455	423
AIC	4470	9936	−688
BIC	4989	−4889	−169

* $p \leqslant .05$.
** $p \leqslant .01$.
*** $p \leqslant .001$.

References

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language, 59*(4), 390–412.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7. <http://CRAN.R-project.org/package=lme4>.

Berthoz, A., Bles, W., Bülthoff, H., Correia Gracio, B., Feenstra, P., Filliard, N., ... Wentink, M. (2013). Motion scaling for high-performance driving simulators. *IEEE Transactions on Human–Machine Systems, 43*(3), 265–276.

Blaauw, G. J. (1982). Driving experience and task demands in simulator and instrumented car: A validation study. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 24*(4), 473–486.

Blana, E. (1996). *Driving simulator validation studies: A literature review*. Institute of Transport Studies, University of Leeds, Working Paper 480.

Brambor, T., Clark, W. R., & Golder, M. (2006). Understanding interaction models: Improving empirical analyses. *Political Analysis, 14*(1), 63–82.

Brickenkamp, R., Schmidt-Atzert, L., & Liepmann, D. (2010). *Test d2-revision. Aufmerksamkeits-und Konzentrationstest*. Göttingen: Hogrefe Testzentrale.

Burnett, G., Irune, A., & Mowforth, A. (2007). Driving simulator sickness and validity: How important is it to use real car cabins? [Special issue] *Advances in Transportation Studies*, 33–42.

Caird, J. K., Willness, C. R., Steel, P., & Scialfa, C. (2008). A meta-analysis of the effects of cell phones on driver performance. *Accident Analysis & Prevention, 40*(4), 1282–1293.

Chiang, D. P., Brooks, A. M., & Weir, D. H. (2001). *An experimental study of destination entry with an example automobile navigation system* (SAE Paper 2001-01-0810). Warrendale, PA: Society of Automotive Engineers.

Classen, S., Bewernitz, M., & Shechtman, O. (2011). Driving simulator sickness: An evidence-based review of the literature. *The American Journal of Occupational Therapy, 65*(2), 179–188.

Engström, J., Johansson, E., & Östlund, J. (2005). Effects of visual and cognitive load in real and simulated motorway driving. *Transportation Research Part F: Traffic Psychology and Behaviour, 8*(2), 97–120.

Feenstra, P., Wentink, M., Correia Grácio, B., & Bles, W. (2009). Effect of simulator motion space on realism in the desdemona simulator. *Paper presented at the driving simulation conference 2009 Europe, Monaco*.

Felnhofer, A., Kothgassner, O. D., Beutl, L., Hlavacs, H., & Kryspin-Exner, I. (2012). Is virtual reality made for men only? Exploring gender differences in the sense of presence. In *Proceedings of the international society on presence research, Philadelphia*.

Felnhofer, A., Hetterle, T., Schmidt, M., Beutl, L., Hlavacs, H., Kryspin-Exner, I., & Kothgassner, O. D. (2014). Challenging gender differences in physical presence: Affective states, spatial abilities and presence in a relaxing virtual park. *Proceedings of the International Society for Presence Research, 2014*, 161–166.

Fischer, M., Eriksson, L., & Oeltze, K. (2012). Evaluation of methods for measuring speed perception in a driving simulator. *Paper presented at the driving simulation conference 2012 Europe, Paris*.

Gemou, M. (2013). Transferability of driver speed and lateral deviation measurable performance from semi-dynamic driving simulator to real traffic conditions. *European Transport Research Review, 5*(4), 217–233.

Godley, S. T., Triggs, T. J., & Fildes, B. N. (2002). Driving simulator validation for speed research. *Accident Analysis & Prevention, 34*(5), 589–600.

Grabe, V., Pretto, P., Giordano, P. R., & Bülthoff, H. H. (2010). Influence of display type on drivers' performance in a motion-based driving simulator. *Paper presented at the driving simulation conference 2010 Europe, Paris*.

Greenberg, J., & Blommer, M. (2011). Physical validity of driving simulators. In D. L. Fisher, M. Rizzo, J. Caird, & J. D. Lee (Eds.), *Handbook of driving simulation for engineering, medicine and psychology* (pp. 7.1–7.22). Boca Raton, FL: CRC Press.

Greenberg, J., Artz, B., & Cathey, L. (2003). The effect of lateral motion cues during simulated driving. *Paper presented at the driving simulation conference North America 2003, Dearborn, Michigan*.

Hill, J. D., & Salzman, J. A. (2012). Enhancing speed perception in virtual environments through training. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 56*, 1772–1776.

Hoffman, L., & Rovine, M. J. (2007). Multilevel models for the experimental psychologist: Foundations and illustrative examples. *Behavior Research Methods, 39*(1), 101–117.

Hoffmann, S., & Buld, S. (2006). *Darstellung und Evaluation eines Trainings zum Fahren in der Fahrsimulation* (Report No. 1960). Düsseldorf: VDI-Verlag.

Hogema, J., Wentink, M., & Bertollini, G. (2012). Effects of yaw motion on driving behaviour, comfort and realism. *Paper presented at the driving simulation conference 2012 Europe, Paris*.

Horrey, W. J., & Wickens, C. D. (2006). Examining the impact of cell phone conversations on driving using meta-analytic techniques. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 48*(1), 196–205.

Hox, J. (2010). *Multilevel analysis: Techniques and applications*. New York: Routledge.

Hurts, K., Angell, L. S., & Perez, M. A. (2011). The distracted driver mechanisms, models, and measurement. *Reviews of Human Factors and Ergonomics, 7*(1), 3–57.

Jamson, H. (2000). Driving simulation validity: Issues of field of view and resolution. *Paper presented at the driving simulation conference 2000 Europe, Paris*.

Jamson, H. (2001). Image characteristics and their effect on driving simulator validity. In *Proceedings of the first international driving symposium on human factors in driver assessment, training and vehicle design* (pp. 190–195). Aspen, CO.

Johnson, D. M. (2007). *Simulator sickness research summary*. Ft. Rucker, AL: U.S. Army Research Institute for the Behavioral and Social Science.

Kappé, B., van Erp, J., & Korteling, J. H. (1999). Effects of head-slaved and peripheral displays on lane-keeping performance and spatial orientation. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 41*(3), 453–466.

Kaptein, N. A., Theeuwes, J., & Van Der Horst, R. (1996). Driving simulator validity: Some considerations. *Transportation Research Record: Journal of the Transportation Research Board, 1550*(1), 30–36.

Kennedy, R. S., & Frank, L. H. (1985). *A review of motion sickness with special reference to simulator sickness* (Report No. NAVTRAEQUIPCEN 81-C-0105-16). Orlando, FL: Naval Training Equipment Center.

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The International Journal of Aviation Psychology, 3*(3), 203–220.

Keshavarz, B., Hecht, H., & Lawson, B. D. (2014). Visually-induced motion sickness: Causes, characteristics, and countermeasures. In K. S. Kale & K. M. Stanney (Eds.), *Handbook of virtual environments* (2nd ed.. *Design implementation, and applications* (2nd ed.), pp. 647–698). New York: CRC Press Taylor & Francis.

King, G., Tomz, M., & Wittenberg, J. (2000). Making the most of statistical analyses: Improving interpretation and presentation. *American Journal of Political Science, 44*, 347–361.

Klüver, M., Herrigel, C., Preuß, S., Schöner, H. P., & Hecht, H. (2015). Comparing the incidence of simulator sickness in five different simulators. *Paper presented at the driving simulation conference 2015, Tübingen, Germany*.

Kolasinski, E. M. (1995). *Simulator sickness in virtual environments* (Report No. 1027). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Science.

Lachlan, K., & Krcmar, M. (2011). Experiencing presence in video games: The role of presence tendencies, game experience, gender, and time spent in play. *Communication Research Reports, 28*(1), 27–31.

Lesch, M. F., & Hancock, P. A. (2004). Driving performance during concurrent cell-phone use: Are drivers aware of their performance decrements? *Accident Analysis & Prevention, 36*(3), 471–480.

Maas, C. J., & Snijders, T. A. (2003). The multilevel approach to repeated measures for complete and incomplete data. *Quality and Quantity, 37*(1), 71–89.

Mayhew, D. R., Simpson, H. M., Wood, K. M., Lonero, L., Clinton, K. M., & Johnson, A. G. (2011). On-road and simulated driving: Concurrent and discriminant validation. *Journal of Safety Research, 42*(4), 267–275.

Miller, J., & Goodson, J. (1960). Motion sickness in a helicopter simulator. *Aerospace Medicine, 31*, 204.

Money, K. (1970). Motion sickness. *Physiological Review, 50*, 1–39.

Mühlbacher, D., Rittger, L., & Maag, C. (2014). Real vs. simulated surrounding traffic – Does it matter? *Paper presented at the driving simulation conference 2014, Paris*.

Mullen, N., Charlton, J., Devlin, A., & Bédard, M. (2011). Simulator validity: Behaviors observed on the simulator and on the road. In D. L. Fisher, M. Rizzo, J. Caird, & J. D. Lee (Eds.), *Handbook of driving simulation for engineering, medicine and psychology* (pp. 13.1–13.18). Boca Raton, FL: CRC Press.

Mullen, N. W., Weaver, B., Riendeau, J. A., Morrison, L. E., & Bédard, M. (2010). Driving performance and susceptibility to simulator sickness: Are they related? *American Journal of Occupational Therapy, 64*(2, SI), 288–295.

National Highway Traffic, Safety Administration. (2012). *Visual-manual NHTSA driver distraction guidelines for in-vehicle electronic devices* (Docket No. NHTSA-2010-0053). Washington, DC: U.S. Department of Transportation, National Highway Traffic Safety Administration

Nicovich, S. G., Boller, G. W., & Cornwell, T. B. (2005). Experienced presence within computer-mediated communications: Initial explorations on the effects of gender with respect to empathy and immersion. *Journal of Computer-Mediated Communication, 10*(2).

Alliance of Automobile Manufacturers (2006). *Statement of principles, criteria and verification procedures on driver interactions with advanced in-vehicle information and communication systems (Rev. June 2006)*.

Osborne, J. W., & Overbay, A. (2004). The power of outliers (and why researchers should always check for them). *Practical Assessment, Research & Evaluation, 9*(6), 1–12.

Pretto, P., Nusseck, H., Teufel, H., & Bülthoff, H. H. (2009). Effect of lateral motion on drivers' performance in the MPI motion simulator. *Paper presented at the driving simulation conference Europe 2009, Monaco*.

Pretto, P., Ogier, M., Bülthoff, H. H., & Bresciani, J.-P. (2009). Influence of the size of the field of view on motion perception. *Computers & Graphics, 33*(2), 139–146.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>.
Ranney, T. A. (2011). Psychological fidelity: Perception of risk. In D. L. Fisher, M. Rizzo, J. Caird, & J. D. Lee (Eds.), *Handbook of driving simulation for engineering, medicine and psychology* (pp. 9.1–9.13). Boca Raton, FL: CRC Press.
Reed, M. P., & Green, P. A. (1999). Comparison of driving performance on-road and in a low-cost simulator using a concurrent telephone dialling task. *Ergonomics, 42*(8), 1015–1037.
Repa, B. S., Leucht, P. M., & Wierwille, W. W. (1982). *The effect of simulator motion on driver performance* (Report No. 820307). Warrendale, PA: Society of Automotive Engineers.
Reymond, G., Kemeny, A., Droulez, J., & Berthoz, A. (2001). Role of lateral acceleration in curve driving: Driver model and experiments on a real vehicle and a driving simulator. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 43*(3), 483–495.
Riby, L., Perfect, T., & Stollery, B. (2004). The effects of age and task domain on dual task performance: A meta-analysis. *European Journal of Cognitive Psychology, 16*(6), 863–891.
Roenker, D. L., Cissell, G. M., Ball, K. K., Wadley, V. G., & Edwards, J. D. (2003). Speed-of-processing and driving simulator training result in improved driving performance. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 45*(2), 218–233.
Rosey, F., & Auberlet, J.-M. (2014). Driving simulator configuration impacts drivers' behavior and control performance: An example with studies of a rural intersection. *Transportation Research Part F: Traffic Psychology and Behaviour, 27*, 99–111.
Roth, P. L. (1994). Missing data: A conceptual review for applied psychologists. *Personnel Psychology, 47*(3), 537–560.
Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review, 103*(3), 403–428.
Schellenbach, M., Lövdén, M., Verrel, J., Krüger, A., & Lindenberger, U. (2010). Adult age differences in familiarization to treadmill walking within virtual environments. *Gait & Posture, 31*(3), 295–299.
Siegler, I., Reymond, G., Kemeny, A., & Berthoz, A. (2001). Sensorimotor integration in a driving simulator: Contributions of motion cueing in elementary driving tasks. *Paper presented at the driving simulation conference 2001, Sophia-Antipolis, France*.
Silverman, D., & Slaughter, R. (1995). *An exploration of simulator sickness in the MH-60g operational flight trainer, an advanced wide field-of-view helicopter trainer* (Report No. AL/HR-TR-1994-0173). Mesa, AZ: Aircrew Training Research Division. Human Resources Directorate.
Singmann, H., & Bolker, B. (2014). afex: Analysis of Factorial Experiments. R package version 0.11-131. <http://CRAN.R-project.org/package=afex>.
Slater, M., McCarthy, J., & Maringelli, F. (1998). The influence of body movement on subjective presence in virtual environments. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 40*(3), 469–477.
Statistisches Bundesamt (2006). *Informations-Technologie in Unternehmen und Haushalten 2005*. Wiesbaden: Statistisches Bundesamt – Pressestelle.
Stoner, A. S., Fisher, D. L., & Mollenhauer, M. (2011). Simulator and scenario factors influencing simulator sickness. In D. L. Fisher, M. Rizzo, J. Caird, & J. D. Lee (Eds.), *Handbook of driving simulation for engineering, medicine and psychology* (pp. 14.1–14.24). Boca Raton, FL: CRC Press.
Stutts, J. C., Reinfurt, D. W, Staplin, L., & Rodgman, E. A (2001). *The role of driver distraction in traffic crashes*. Washington, DC: AAA Foundation for Traffic Safety.
Törnros, J. (1998). Driving behaviour in a real and a simulated road tunnel – A validation study. *Accident Analysis & Prevention, 30*(4), 497–503.
Uliano, K., Lambert, E., Kennedy, R., & Sheppard, D. (1986). *The effects of asynchronous visual delays on simulator flight performance and the development of simulator sickness symptomatology* (Report No. NAVTRASYSCEN 85-D-0026-1). Orlando, FL: Naval Training Systems Center.
Wang, Y., Mehler, B., Reimer, B., Lammers, V., D'Ambrosio, L. A., & Coughlin, J. F. (2010). The validity of driving simulation for assessing differences between in-vehicle informational interfaces: A comparison with field testing. *Ergonomics, 53*(3), 404–420.
Warner, H. D. (1993). *Flight simulator-induced sickness and visual displays evaluation* (Report No. AL/HR-TR-1993-0056). Williams Air Force Base, AZ: Aircrew Training Research Division.
Wierwille, W. W., Casali, J. G., & Repa, B. S. (1983). Driver steering reaction time to abrupt-onset crosswinds, as measured in a moving-base driving simulator. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 25*(1), 103–116.
Yan, X., Abdel-Aty, M., Radwan, E., Wang, X., & Chilakapati, P. (2008). Validating a driving simulator using surrogate safety measures. *Accident Analysis & Prevention, 40*(1), 274.
Young, K., & Regan, M. (2007). Driver distraction: A review of the literature. In I. J. Faulks, M. Regan, M. Stevenson, J. Brown, A. Porter & J. D. Irwin (Eds.), *Distracted driving* (pp. 379-405). Sydney, NSW: Australasian College of Road Safety.
Zeeb, E. (2010). Daimler's new full-scale, high-dynamic driving simulator – A technical overview. *Paper presented at the driving simulation conference 2010, Paris*.