Please do not remove this page

Systematic review of driving simulator validation studies

Wynne, Rachael; Beanland, Vanessa; Salmon, Paul M

https://research.usc.edu.au/esploro/outputs/journalArticle/Systematic-review-of-driving-simulator-validation/99451504402621/filesAndLinks?index=0

Wynne, R., Beanland, V., & Salmon, P. M. (2019). Systematic review of driving simulator validation studies. Safety Science, 117, 138–151. https://doi.org/10.1016/j.ssci.2019.04.004
Document Type: Accepted Version

Link to Published Version: https://doi.org/10.1016/j.ssci.2019.04.004

UniSC Research Bank: https://research.usc.edu.au
research-repository@usc.edu.au
CC BY-NC-ND V4.0
Copyright © 2019. This manuscript version is made available under the CC-BY-NC-ND 4.0 license
http://creativecommons.org/licenses/by-nc-nd/4.0/
Downloaded On 2024/05/02 09:25:21 +1000

Please do not remove this page

Systematic Review of Driving Simulator Validation Studies

Rachael A. Wynne [a*], Vanessa Beanland [a, b], and Paul M. Salmon [a]

[a] Centre for Human Factors and Sociotechnical Systems, University of the Sunshine Coast, Sippy Downs, Australia

[b] Department of Psychology, University of Otago, Dunedin, New Zealand

[*]corresponding author

Author details

Rachael A. Wynne, Locked Bag 4, Maroochydore DC, QLD, 4558

rachael.wynne@research.usc.edu.au, ORCiD: 0000-0003-2623-1141

Vanessa Beanland, vanessa.beanland@otago.ac.nz, ORCiD: 0000-0002-7961-4114

Paul M. Salmon, psalmon@usc.edu.au, ORCiD: 0000-0001-5534-9830

Abstract

Driving simulators are a common tool for researching driver behaviour, providing practical, safe, and controlled environments. Despite their frequent use in research, there is relatively little evidence confirming their validity (i.e., how accurately they represent or reproduce real-world driving). Moreover, there is inconsistency in both the types of simulators used, and the operationalisation of "real-world" driving in validations. This systematic review was undertaken to evaluate the evidence regarding driving simulator accuracy when compared with real-world driving. The review included 44 studies reporting a direct comparison between simulated driving and on-road driving in a vehicle. Measures reported for comparison varied but included mean speed, speed variability, lateral position, overall driving performance, and number of driving errors. Simulators in approximately half of the studies achieved absolute or relative validity, whereas one third produced non-valid results. To understand this further, the fidelity of simulators was considered, however this further clouded our understanding as the relationship between simulator fidelity and validity was not straightforward. The findings suggest that the reporting of driving simulator studies requires improvement, particularly around the validation evidence associated with the simulator, the specific details of the simulated driving environment, and the outputs of statistical analyses. Guidelines are proposed for future research to ensure consistency in the conduct, and reporting, of simulator-based research.

Keywords: driving simulator; validation; fidelity; real-world driving

Systematic Review of Driving Simulator Validation Studies

1. Introduction

Driving simulators allow researchers to examine complex behaviours in a controlled environment that might otherwise not be practical, safe, or ethical (Calhoun and Pearlson, 2012). Simulators first emerged in the 1930s (Lauer, 1960), and have been commonly used by researchers to investigate a range of driver behaviours, including the effects of technologies, devices, and road infrastructure, ranging from variable message signs (e.g., Comte and Jamson, 2000) and in-vehicle systems (e.g., Abe and Richardson, 2005; Lin et al., 2009) to mobile phone use (Choudhary and Velaga, 2019) and automated vehicles (e.g., Eriksson and Stanton, 2017).

An issue with any laboratory-based experiment is the validity and reliability of the apparatus; that is, the extent to which they accurately and consistently emulate real-world performance. *Reliability* refers to the ability of a simulator to report consistent results over time. Within driving research, reliability studies may involve participants completing the simulator task multiple times, with analyses comparing driving performance across time (e.g., Davenne et al., 2012). *Validity* refers to the ability of a simulator to accurately represent real-world driving. There are various forms of validity (Annett, 2002; Stanton, 2016), with most studies assessing two: absolute validity and relative validity (Blaauw, 1982).

Absolute validity occurs when the values obtained in a simulator (e.g., speed or lateral position) match those obtained in a real vehicle in absolute terms. Establishing absolute validity requires direct comparison of simulated and real-world driving, with statistical tests (e.g. *t*-tests) showing no significant difference between the values for the two types of driving. Even though absolute validity is always desirable it is not always achieved, and in some contexts relative validity may be acceptable.

Relative validity occurs when simulator results show the same patterns or effects as real-world driving. Empirical tests of relative validity could take different forms, depending on the study design. For example, a study could employ a 2×2 design with drive type (simulator vs. real-world) and distraction (distraction-free vs. phone use) as factors. In this design, relative validity would be demonstrated if there was a main effect of distraction (e.g., mean speed reduced when distracted in both drive types) even if the study failed to establish absolute validity (i.e., mean speeds differed between simulated and real-world driving). In a simpler design, relative validity could be demonstrated if simulated and real-world driving measures were positively correlated, but differed in magnitude (i.e., drivers who exceeded the speed limit in the real world also drove faster in the simulator, but at different absolute speeds).

Although there is a vast body of driving simulator-based research, and researchers acknowledge the importance of simulator validity, there are comparatively few studies that have explicitly sought to validate a driving simulator by directly comparing simulated and real driving. Moreover, there is no standardised method for assessing simulator validity, and so there is considerable variation in validation study methodologies. The present systematic review was designed to assess driving simulator validation studies published in the peer-reviewed literature to understand how accurately driving simulators represent or reproduce performance equivalent to "real-world driving". Though there have been previous reviews addressing the question of simulator validity (e.g., Blana, 1996; Mullen et al., 2012), these papers obviously omit recent literature, which is ever-increasing. Moreover, to our knowledge previous reviews have not conducted systematic literature searches and therefore have not captured all relevant papers.

1.1. Progression of driving simulators

In the last 40 years technological advancements have increasingly enabled higher quality computer processing and graphics along with more sophisticated and accurate control devices. Most simulators are now dynamic, with the actions of the driver resulting in changes in the driving environment. The visual quality of the virtual environment has also increased, with current simulators able to include elements such as controllable traffic, different road users (vehicles, motorcycles, bicycles, pedestrians), and interactive modifiable features such as billboards and railway level crossings. These elements can be programmed to modulate in response to the driver's actions or as a pattern that the driver must respond to (e.g., the integration of traffic simulation modelling into the driving simulator; Jeihani et al., 2017). There have also been marked improvements in the physical equipment used. The most basic simulator configurations have participants seated in a single chair, using a limited movement steering wheel or joystick, but it is becoming increasingly common to see simulators incorporating the use of a full or partial vehicle body and motion platform (Bouchner, 2016). In summary, technological changes have produced simulators that better resemble real driving in terms of vehicle controls and the visual environment. Although this level of realism is thought to play an important role in validity (Greenberg and Blommer, 2012), there is limited empirical research that directly tests the extent to which this matters.

1.2. Simulator fidelity

Technological advances have decreased production costs, leading to an increase in the number of driving simulators, but this has led to a wide variability in simulator design. A key issue for driving simulators is fidelity; the extent to which they emulate driving in the real-world. Kaptein et al. (1996) categorised simulator fidelity based on physical design elements, such as the inclusion of a motion base and vehicular controls. According to their criteria, a "high-level" driving simulator would provide close to a 360° field of view projected onto large screens, full-feedback motion base, and have participants seated in a vehicular cab with full controls (Kaptein et al., 1996). In comparison, a "lowlevel" simulator may employ a single computer monitor with simplistic controls, such as a keyboard, to direct the vehicle through a scene, mimicking driving (Kaptein et al., 1996). Despite being more affordable, these simulators do not provide the same level of physical realism and therefore some argue that the findings from those simulators cannot be regarded as equivalent to those from higher physical fidelity simulators (Caird and Horrey, 2011). Some researchers use the term "physical validity" when describing physical fidelity (e.g., Mullen et al., 2012); however, here we prefer the term fidelity to avoid confusion with other aspects of validity, because fidelity relates to research tools and independent variables (e.g., characteristics of the simulator and tasks used) whereas absolute and relative validity relate to research outcomes.

Although most descriptions of simulator fidelity emphasise the fidelity of the physical construction or hardware, there are many other domains that can be considered including: behavioural fidelity, task fidelity, functional fidelity, perceptual fidelity, motion fidelity, objective fidelity, psychological fidelity, and concrete fidelity (de Winter et al., 2007; Goode et al., 2013). With many dimensions of fidelity to consider, it is difficult to quantify the overall fidelity of a simulator, as it may be high on one aspect of fidelity (e.g., vehicle controls) but not another (e.g., the visual driving environment). Indeed, there is currently no common approach to classifying simulator fidelity, with "low", "medium", and "high" classifications used without a standardised measure.

For the purpose of this review, simulation fidelity will be considered primarily in terms of vehicle controls (physical), field of view (visual), and the kinaesthetic feedback (motion) provided to the driver, which are the measures considered most crucial to driving performance (Goode et al., 2013). Ideally analyses of simulator fidelity should also address the fidelity of the software and/or visual environment that is used to develop the driving scenario. Unfortunately, this could not be feasibly assessed in the current review. Even when researchers publish images of their simulation scenarios (e.g., Polders et al., 2015; Yan et al., 2008) they are static images and do not fully represent the participant's visual experience. Additionally, there are many factors upon which the quality of the visual environment would vary, such as: hardware capacity, programmer skill, environmental complexity (e.g., rural road vs city street), the number of other road users, and the specific driving tasks (e.g., straight road driving, gap acceptance, merging). Thus, even for two simulators using identical software and hardware, the realism of the visual environment could differ.

1.3. Measures of real-world driving

When comparing simulators and real-world driving, simulator fidelity is only one half of the equation. The other factor that impacts the comparison of simulated and real driving is the operational definition of "real-world" driving. Several approaches are evident in the literature, including self-reported driving behaviour (e.g., Ba et al., 2016; Szlyk et al., 1992), allied health assessments (e.g., Lauridsen et al., 2016; van Wolffelaar et al., 1988), and on-road drives in instrumented vehicles (e.g., Helland et al., 2016; Lauer and Suhr, 1958). These techniques use diverse variables to compare driving performance; involving both physical (e.g., lane deviation, speed, reaction time, crashes) and cognitive (e.g., cognitive load, divided attention, situation awareness) driving behaviours (Caird and Horrey, 2011). To present the most meaningful comparisons between simulated and real driving, the current study focused on studies that incorporated objective measures from on-road drives as the measure of real-world driving.

1.4. Objectives

The primary objective of this review is to evaluate evidence on the validity of driving simulators. Secondary objectives of this systematic review are to:

- determine the range of real-world driving measures used within driving simulator research (e.g., on-road driving, self-report questionnaires, neurological assessments) and identify the prevalence of direct comparisons with on-road driving;
- identify the suite of dependent measures used for comparisons (e.g., mean speed, lane position, errors) and to establish the validity of those variables in a simulator; and
- evaluate whether, and how, simulator fidelity impacts the validity of driving simulators.

2. *Methods*

2.1. Protocol

The systematic review was conducted in accordance with the PRISMA guidelines (Moher et al., 2009). The review protocol is outlined in the following subsections. Initially a meta-analysis was considered, but after collating the studies it became clear this was not feasible because of inconsistencies in results reported across studies. Specifically, studies differed in the measures reported, meaning few had directly comparable results. In addition, descriptive statistics were not commonly reported and in some cases no statistics were reported (i.e., researchers claimed there was a significant correlation, or no significant difference, without reporting data or analyses). Finally, even if descriptive statistics and full statistical results were reported, there may be differences in how summary measures were derived. For example, "mean speed" and "speed variation" could be calculated across a short straight segment or an entire naturalistic route, including curves and intersections (which would be expected to show greater variability). It would not be possible to draw meaningful conclusions between two studies with differing approaches to the calculation of these measures.

2.2. Information sources

Four databases were searched: PsycINFO, PubMed, ScienceDirect, and the Transportation Research Information Database (TRID). This allowed the authors to review both published (PsycINFO, PubMed, ScienceDirect) and unpublished or "grey" literature (indexed by TRID). The initial search was conducted on 8-10 August 2016, with two follow-up searches (12 July 2017, 1 November 2017) to identify additional publications. Items 2019 before 31 October 2017 were included.

2.3. Search strategy and study selection

The following Boolean search terms were generated based on an initial literature search and were designed to gather all relevant derivations of the base terms:

(1) "driv* OR vehicle OR automobile OR car OR truck OR lorry OR van OR motorcycl*" AND
(2) "simulat*" AND
(3) "valid* OR evaluat* OR compar*"

The first search identified 23,115 items (see Table 1). Figure 1 shows the selection process used to exclude out-of-scope documents. Duplicate items were identified using

EndNote and removed (*n*=1,478). Following the initial search, most articles were rejected through review of title and abstract (*n*=21,312), predominantly for being unrelated to driving (e.g., pharmacology or medicine focused). Items were also excluded if they did not include original results or were a systematic review or meta-analysis.1 Item selection was performed by the first author, and the second author performed a random, independent review of 10% the search results, to ensure no items were incorrectly excluded. Only two items (0.09%) were identified as potentially incorrect exclusions via review of title and abstract; however, subsequent review of the full article confirmed that they did not meet the inclusion criteria (outlined in Table 2). The second author reviewed a sample of the full papers at the time of data review, and provided consultation when results were being analysed.

Articles using motorcycle simulators were removed because of fundamental differences in terms of their physical properties; thus, the technical specifications required to achieve high fidelity simulation are quite different as motorcyclists experience much greater range of motion. Bicycles were excluded as they are not a motor vehicle. Trucks and lorries were also removed as they represent a different group of drivers (i.e., professional drivers) and therefore may have presented different results.

After screening, 325 items were evaluated based on methods and results. Studies were retained if they reported a comparison of a driving performance measure assessed both in the simulator and the "real-world". At this stage, 166 items were identified for potential inclusion. These included items using an on-road drive (*n*=80); accident or driving history supported by insurance or police records (n=4); neuropsychological assessment (*n*=36); self-report (*n*=50); and simulator training program (*n*=11).

Two follow-up searches with the same parameters were conducted 11 and 15 months after the initial search to identify subsequent publications. The first follow-up resulted in three additional items satisfying all inclusion criteria that were subsequently included in the final review. In the final search 11 additional items were identified, with one eventually included following application of the exclusion criteria.

The current analysis focused on those that incorporated an on-road drive in a vehicle as the measure of real world driving only. Some studies were subsequently excluded from the review as the analysis reported another aspect of behaviour, such as sleepiness (Filtness et al., 2014) or blood alcohol concentration (Allen et al., 1978), rather than driving performance. Further items were removed after examining the methodology and results; most as they omitted a direct comparison between the simulator and on-road drives. This resulted in 44 items being included in the analysis, and notable exclusions can be seen in Table 3.

2.4. Bias

Within systematic reviews, there is an inherent risk of bias. The current review attempted to minimise biases wherever possible. Publication bias is a common issue, as certain results may be deemed “non-publishable” by reviewers or editors. To minimise the impact of publication bias, we searched the TRID database for “grey” literature, including unpublished papers and research reports. However, another source of bias stems from those conducting the original research. That is, researchers may selectively report measures or even entire studies, such as omitting results that are inconsistent with their hypotheses or research aims. In this case, it is not possible to remove this source of bias, instead one must be aware that there may be additional research that has not reached the point of dissemination which may show further support for the non-validity of simulators. It is still important that all research, be it significant or non-significant results, be communicated to the wider audience as it provides opportunities to improve our current research understanding and practice. In the context of simulator validation studies, it is most plausible that researchers would fail to report measures or studies that suggested a simulator was invalid; thus, review results may be biased towards over-estimating the validity of simulators. Finally, within any literature review there is a risk of bias on behalf of those conducting the review. To address this bias, strict eligibility criteria were established prior to the commencement of the review (see Table 2), and a second analyst was used to establish reliability.

2.5. Simulator fidelity scoring

A unified approach for quantifying fidelity was required to facilitate comparisons between the various simulators. Each simulator was rated on three fidelity measures: visual, motion, and physicality (see Table 4). Each measure was scored on a 5-point scale, so the range of possible total scores was 3–15 with higher scores representing greater fidelity. Characteristics of a low-fidelity simulator would be a fixed-base, single computer screen with video game steering wheel (e.g., Davenne et al., 2012). Medium-fidelity simulators incorporate multiple screens with a limited field of view (FOV; i.e. 135°), partial motion platform, with an arcade-style driver seat, steering wheel, pedals, and controls. High-fidelity simulators include features such as a 270–360° FOV with forward and rear projections, full motion platforms with road sensation feedback, and a full vehicular cabin (e.g., Fildes et al., 1997).

Ratings were based on the details reported in the articles. In two cases insufficient detail was provided (Freund et al., 2002; Galski et al., 1992), so the fidelity score was based on standard specifications for the simulator brand and model described. One author scored all 52 simulators in the review, with a second rater independently scoring 16 (31%). Inter-rater agreement was 81.3% with a strong correlation between scores (r=0.96). Cohen's kappa, a measure of agreement whereby agreement due to chance is removed, was κ=0.79, indicating substantial agreement between raters (Kaplan and Saccuzzo, 2009).

3. Results

3.1. Reviewed studies

The database search yielded 44 items (i.e., articles, papers, or reports) that met inclusion criteria (see Supplementary Materials for full details). All items were originally published in English between 1977 and 2017. Studies incorporated an experimental or quasi-experimental approach, in which drive type (simulator vs. real) was manipulated within- (n=30) or between-subjects (n=13), with one study using a mixed within-between design. Some studies also involved other manipulations which were outside the scope of this review, such as comparing levels of driving experience (Carter and Laya, 1998) or cognitive impairment (Freund et al., 2002).

Where sample size could be determined, participant numbers ranged from 8 to more than 2000. It was not always possible to identify the number of participants as some naturalistic observations were conducted in the field using external recording devices, such as radar speed recorders. Participant gender was not consistently reported, especially studies involving naturalistic observations of unidentified drivers. Reported ages of participants varied from 16 to 88 years.

Methods used to gather data for on-road drives included: field observations with laser or radar speed controls, field recordings of intersections or road sections, instrumented vehicles, and recordings of participants' own vehicles.

3.2. Simulator fidelity

Most items used a single simulator, but some papers compared multiple simulators (Edwards et al., 1977; Lee et al., 2013) and in other cases the same simulator was used in multiple papers (e.g., Lee et al., 2013; Mueller, 2015). Overall, 52 simulators were used, which varied in design including desktop PC-based simulators, arcade style singles-eat simulators, full vehicles placed on motion platforms, and custom designs.

Fidelity scores ranged from 4 to 15 (M=9.7, SD=3.2). Nearly half the simulators rated (n=24) received intermediate scores (8–10) and therefore were considered medium fidelity simulators. The other half were evenly split between low fidelity (score 4–7; *n*=14) and high-fidelity (score 12–15; *n*=14) simulators. The relationship between simulator fidelity and validity was not as clear as one would expect, that is not all high-fidelity simulators were able to achieve validity, and as will become clearer below some simulators were able to achieve validity on some measures while not on others. With this in mind, approximately half of all simulators achieved either absolute or relative validity.

3.3. Summary of measures

Several measures were reported by multiple researchers, but there was no single common variable across all studies. Mean speed and/or speed variation have been suggested as the most reliable variable to determine simulator validity, as speed has repeatedly demonstrated absolute validity (e.g., Bella, 2008; Blaauw, 1982; Mueller, 2015; Reed and Green, 1999). Mueller (2015) also noted lateral deviation or lane positioning as demonstrating strong relative validity. Additional variables repeatedly reported included driving errors (e.g., Edwards et al., 1977; Freund et al., 2002; Galski et al., 1992), and global measures of driving performance (e.g., Lee et al., 2003; Lew et al., 2005). Finally, other popular measures included direction of eye movements/gaze, (e.g., Carter and Laya, 1998), braking behaviours (e.g., Zoller et al., 2019) and steering (e.g., Reed and Green, 1999). The validity of these measures was also evaluated as they address other relevant driving behaviours which may impact on speed, lane position, and errors.

Each measure reported in each study was assessed to determine whether the presented data suggested equivalence between the simulator and real-world driving. Specifically, we considered whether the comparison suggested absolute, relative or no validity, based on the data and statistics reported. Statistical significance was assessed using $\alpha=0.05$, unless stated otherwise. Comparisons were classified as establishing absolute validity if measures from the simulated and real drives were directly compared using appropriate inferential statistics, such as *t*-tests or ANOVA, and the results showed no significant difference between drives. In contrast, if the data analysis focused on comparing trends between the simulator and real-world drives, it would not be possible to demonstrate absolute validity but could be possible to demonstrate relative validity, for example if measures for the two drives were significantly correlated. Finally, results were considered “not valid” if data analysis showed a significant difference between drives, or a non-significant correlation.

3.4. Speed or speed variation

Speed or speed variation was the most common dependent measure, reported in 21 of 44 items. Most items found speed or speed variation was equivalent between the simulator and real drive (i.e., either a significant correlation or non-significant difference between drives), thus achieving some form of validity (see Table 5). Most studies achieving validity used high-fidelity simulators, according to the fidelity rating scale (Abdel-Aty et al., 2006; Branzi et al., 2017; Fildes et al., 1997; Lee et al., 2013). Only two studies achieved validity with a low-fidelity simulator (Blaauw, 1982; Lee et al., 2013).

Of the 13 studies that achieved valid results, nine reported absolute validity and six relative validity; with two reporting both absolute and relative validity. Low-fidelity simulators achieved relative validity only, whereas medium- and high-fidelity simulators reported both absolute and relative validity.

Ten studies reported non-valid speed comparisons (see Table 5), of which three used a low-fidelity simulator and four used medium-fidelity. Three studies using high-fidelity simulators reported significant differences between simulator and on-road driving, but the nature of these differences was inconsistent across studies, with two studies reporting higher speeds in the simulator (Hallvig et al., 2013; Senserrick et al., 2007) and the third higher on-road (Fors et al., 2013). Senserrick et al. (2007) also reported significantly greater speed variation in the simulator.

Two studies, both using medium-fidelity simulators, provided mixed validation support. Alm (1996) found no significant difference for mean speed (suggesting absolute validity), but a significant difference for speed variation, with greater variation in the simulator than on-road. This significant difference was found with the motion base both on and off. While Alm did not propose an explanation, others have suggested this difference is due to the reconstruction of the virtual environment, traffic level, and the perceived lack of risk in the simulated environment (Bella, 2008). Similarly, McAvoy et al. (2007) found inconsistent results when comparing simulated and real driving through work zones, with different patterns of results for control and treatment sites. Speeds were significantly faster in the simulator at the start and end of the control site, and the middle of the test site, but equivalent in other sections. When exploring the results further, McAvoy et al. identified that drivers maintained a constant speed in the simulator but reduced speed over time in the field. The inconsistences in speed were attributed to the replication of the virtual environment, specifically the retro-reflectivity and lumens of hazard lights, and perceived risk in the real world being insufficiently replicated in the simulator.

3.5. Lane position or variation in lane position

Lateral position measures (e.g., lane position, standard deviation of lane position) were the next most commonly reported driving measures; outlined in Table 6.

Only four studies achieved validity for lateral position, and all involved medium-fidelity simulators. These were a mix of relative and absolute validity (see Table 6). Alm (1996) again reported both absolute validity and no validation support on different aspects of these measures. While no significant difference was found for mean lateral position, a significant difference was found for the variation in lateral position. When the simulator motion base was off, drivers varied their position more than in the real world and with the motion base on.

Nine studies reported non-valid results (i.e., significant differences measured between the simulator and on-road), with findings from two high-fidelity, five medium-fidelity, and two low-fidelity simulators (see Table 6). Of those reporting non-valid results, most found more variation in lateral position in simulators compared with real driving (Blaauw, 1982; Daurat et al., 2013; Hallvig et al., 2013; Helland et al., 2013; Wang et al., 2010). Two studies focused on mean lane position, found participants drove closer to the centre line (Harms, 1996; Wade and Hammond, 1998) and closer to the tunnel wall (Tornros, 1998) in the simulator than in the real world. It was proposed that these differences were again due to the perception of risk in the simulated environment (Tornros, 1998) and the lack of haptic feedback in a low-fidelity simulator (Wade and Hammond, 1998).

3.6. Line crossings and lane change behaviour

Line crossings, specifically inappropriate line crossings, were compared in four items (see Table 7). Absolute validity was found in one medium-fidelity and one high-fidelity simulator. However, for low-fidelity simulators, significantly more crossings occurred in the simulator than on-road. Both studies attributed this to the inertia of the vehicle in real driving, such that it tempers the effects of lateral deviation, whereas in simulated driving it is amplified, and it is more difficult to stay in the designated lane (Davenne et al., 2012; Philip et al., 2005). Davenne et al. (2012) also acknowledged that the perceived consequence of rule violation in the simulator is almost non-existent, and when combined with the lower perceived risk, there is less incentive to stay in one’s lane.

Yun et al. (2017) studied lane changing behaviours in the high-fidelity simulator and achieved absolute validity between the real and simulated drives.

3.7. Overall driving performance and errors

A trend common in the allied health sector is to have an observer rate driving in terms of an overall score, or number of errors, particularly when determining fitness to drive. Ten studies adopted this approach, and seven found comparable performance in the simulator and on-road (see Table 8). Of those seven studies, six used low-fidelity simulators with one medium-fidelity simulator, two studies reported absolute validity (Barker et al., 1978; Meuleners and Fraser, 2015) with the remainder reporting only relative validity.

The three studies that found non-valid differences all employed a medium-fidelity simulator. Results were inconsistent across these studies. Edwards et al. (1977) did not report specific statistics, stating only that there were no significant correlations between simulated driving and on-road performance. Hulme and Thorpe (2013) found significantly more driving errors in the simulator, whereas Shechtman et al. (2009) found more in the on-road drive. Analysis of driving errors found that only those that related to speed, vehicle position, and signalling were significantly higher in the simulator than on-road (Shechtman et al., 2009).

3.8. Other measures

Eleven studies reported findings across eight other measures of driving performance, including braking behaviour (Lee et al., 2002; Zoller et al., 2019), travel time (Hou et al., 2014; Johnson et al., 2011), and eye fixations (Carter and Laya, 1998; Fors et al., 2013; Mueller, 2015). Some studies reported multiple measures, and overall six studies achieved simulator validation across six measures, whereas six studies found non-valid results across five measures, some of which were found to be valid in other studies (see Table 9). Those that reported validity used medium- and high-fidelity simulators only, while non-valid results were reported across all levels of simulator fidelity.

For braking behaviour, Lee et al. (2002) found similar behaviours in a high-fidelity simulator (relative validity), whereas Zoller et al. (2019) used low- and medium-fidelity simulators and found drivers would brake significantly earlier when on-road than in all forms of simulator.

Johnson et al. (2011) and Li et al. (2013) both examined drivers' heart rate while driving on-road and in a medium-fidelity simulator. Li and colleagues achieved absolute validity with no significant differences, yet drivers in Johnson et al. found a significantly higher maximum heart rate when on-road than in the simulator. The higher heart rate was attributed to the greater workload and psychological stress associated with driving in the real-world.

Steering was examined by Mueller (2015) in a high-fidelity simulator and by Reed and Green (1999) in a medium-fidelity simulator. Reed and Green found a moderate positive correlation between on-road and simulator for steering behaviours, providing relative validity. Conversely, Mueller found greater steering variations and higher reversal frequency when in the simulator.

Three measures achieved only absolute validity; beta waves measured using an EEG (Li et al., 2013), travel time (Hou et al., 2014; Johnson et al., 2011) and vehicle headway (Risto and Martens, 2014). All used medium-fidelity simulators except for Hou et al. (2014) who employed a high-fidelity simulator.

Studies of mental workload and eye fixations reported only non-valid results. Carter and Laya (1998) used a low-fidelity simulator and found significantly more fixations per second in a significantly smaller range of zones, with more time being spent in fixation in the simulator than on-road. Fors et al. (2013) reported similar findings, with significantly more road gaze fixation in the simulator than on-road, with a smaller radius of gaze, so the gaze pattern was more centralised on the road ahead. Participants in Mueller's (2015) study appeared to incorporate greater horizontal and vertical gaze dispersion in the high-fidelity simulator than on-road. Mueller also assessed mental workload, finding that the simulator required higher levels of mental effort, but resulted in lower performance overall than the real drive. This supports the earlier work of Alm (1996) who found that the simulator was significantly more physically and mentally demanding than the on-road drive.

4. Discussion

Given the prevalent use of simulation in road safety research, studies that attempt to assess the validity of driving simulators are critical. The current review was conducted to evaluate evidence regarding the validity of driving simulators when compared with real-world driving. The review revealed that there have been relatively few validation studies, and that the methods have varied considerably, as described further in the following subsections.

4.1. Main findings

There was little consistency in the dependent measures used to assess differences between the simulator and on-road drive. Of the 80 items initially identified as potentially within scope, nearly half were removed for either not including a driving measure or not reporting a direct comparison between the simulator and on-road drive. In the remaining studies several dependent measures were used, predominantly speed, speed variation, lateral position, errors and/or global driving performance. One-quarter of the studies also reported other measures such as eye fixation (e.g., Carter and Laya, 1998), steering (e.g., Mueller, 2015), and mental workload (e.g., Alm, 1996). Although some items reported multiple measures, there was no single measure reported by all studies, making it impossible to compare findings reliably. This indicates the need to develop proposals for consistent future research practice.

Given that there was no single variable that all studies addressed, it was not possible to formally evaluate all studies on the same outcome measure. However, half of the 44 items evaluated reported equivalence between the real-drive and the simulator, that is, either absolute or relative validity. One third of the studies reported significant differences and/or non-significant correlations between driving conditions, suggesting that the simulators did not provide valid representations of real-world driving performance. The remaining studies presented both valid and invalid results, suggesting the simulators in question were not universally valid. Moreover, this suggests that validity on one measure in one simulator does not guarantee that other measures will be valid in the same simulator, or that the same measure will be valid in other simulators. This in turn could call into question the relevance of some simulator-based research findings, particularly where research is used to inform government policy, training programs, and legislation. If researchers are not using valid simulators, or simulators that are only valid for selected measures, conclusions may be incorrect or biased.

4.2. Simulator fidelity

An aim of the review was to assess the impact of simulator fidelity on the research outcomes and validity of driving simulators. The findings suggest that the relationship between fidelity and validity was not straightforward. Some might assume low-fidelity simulators would be the least likely to return valid results, being less indicative of the real-world, when in fact some low-fidelity simulators achieved validity (e.g., Lee et al., 2013; Mayhew et al., 2011) and some high-fidelity simulators invalid findings (e.g., Hallvig et al., 2013; Senserrick et al., 2007). To explore this further, we can look to the findings of line crossings and braking behaviour, where the fidelity of the simulator consistently impacted the findings.

In the case of line crossings, medium- and high-fidelity simulators were found to be valid but low-fidelity simulators were not. This can be understood by examining differences between the simulators. Both low-fidelity simulators incorporated single computer screen visual environments (Davenne et al., 2012; Philip et al., 2005), whereas the high- and medium-fidelity simulator provided a wider field of view (Daurat et al., 2013; Fors et al., 2013); one possibility is that when the driver can see more of the visual environment they can remain in their lane more easily.

Physical fidelity may also be influential; high physical similarity to real driving may assist in maintaining lane position and using the visual cues they would normally rely on in the real-world. Recent research investigated this by manipulating the presence of a vehicular cabin in what would be considered a low-to-medium fidelity simulation (Mecheri and Lobjois, 2018). There was a significant effect of the physical environment on lateral deviation, indicating that the cabin allowed drivers to better perceive and maintain appropriate lane positioning.

Within several items considered for review (Daurat et al., 2013; Davenne et al., 2012; Fors et al., 2013; Philip et al., 2005) researchers did not report the direction of observed differences (i.e., whether drivers were more likely to cross towards or away from the centre line). This would be interesting to note when considering whether perception of risk, such as from oncoming traffic, was a factor in this behaviour. It is important to note that this only applies to line crossings, as there were still significant differences in the variation of lateral position in medium- and high-fidelity simulators.

4.3. Notable findings

An interesting finding was that of Shectman et al. (2009), who analysed the types of errors that were committed by drivers in the simulator and on road. While some significant differences existed between the types of errors, it was only those that related to speed, vehicle position, and signalling that were significant and they were significantly higher in the simulator than real-drive (Shectman et al., 2009). This aligns with the studies that found differences in speed and lane position between the simulator and real world (see Tables 5 and 6). While Shectman et al. did not propose any suggestions for this increase in errors, it is commonly understood that the level of perceived risk is lower in simulators (Ranney, 2012). Therefore, drivers may, despite opposing instructions, engage in riskier behaviours, or feel there are minimal consequences for noncompliance with rules when in the simulator, and thus have more errors.

Another interesting result was the apparent contradiction between mental workload and heart rate in the simulator. Alm (1996) and Mueller (2015) both found that participants' self-reported mental effort in the simulator was higher than that of the real drive, which can be accounted for when considering the perceptual cycle model (Neisser, 1976). In accordance with this model, as one develops experience with a particular action or situation, the workload required to engage with that situation is reduced by activated schemata which direct perceptual exploration and action. As drivers have well developed schemata for real world driving but not for simulator driving, schemata for real environments cannot be applied unless the simulated environment is highly indicative of the real-world. Thus, more mental effort is required to complete the drive, and performance is also likely to be lower as the experience of a real vehicle may not transfer.

As the apparent mental effort required is greater, and heart rate increases with mental load (Wilson, 2002), it would be expected that the heart rate of participants would be higher in the simulator than in the real-world. Yet, Johnson et al. (2011) found that drivers' heart rate was lower in the simulator and Li et al. (2013) found no significant difference; both used medium-fidelity simulators. These findings may seem inconsistent but could be explained by the fact that perceived risk is lower in simulators than in the real-world (e.g., Bella, 2008; McAvoy et al., 2007). Specifically, drivers may experience heightened stress in the real-world, not due to mental workload but the perception of danger, which would account for elevations in heart rate from a baseline (Johnson et al., 2011).

4.4. Implications

Some of the findings are alarming given that driving simulator studies are widely used to inform road safety policy and practice. Of particular concern is the fact that only half of the driving simulators were found to be valid and some were valid for one measure but not others. It is notable that articles reporting simulator studies do not often urge caution when interpreting study findings. A clear implication is that future articles should include such statements, particularly if the validity of the simulator is not established. Ideally this would see authors report empirical validation evidence for their own simulator, and not relying on other simulators as support for validity. Even if modelled on a previously validated simulator, each set-up is unique and should be validated for those specifications. If a validation study is not feasible, authors should emphasise to those who intend to draw conclusions from their results that there is a possibility it may not be valid.

A further point to consider is that some items included other measures, including some beyond the realm of standard driving related measures, such as eye movements (Carter and Laya, 1998; Fors et al., 2013; Mueller, 2015) and mental workload (Alm, 1996). These atypical measures will also need to be considered in future validation studies, as their inclusion in many simulator studies implies that researchers believe the cognitive resources necessary to complete a drive in a simulated environment are the same as those of a real drive.

4.5. Limitations

The current review is not without limitations, and therefore the findings must be interpreted with caution. Many of these limitations are data-driven; that is, they arose because limited or inconsistent information was reported in the examined research literature.

There was no consideration of the software or visual environment when comparing the simulators used on the research. As noted, this information could not be appropriately identified in source material. Nevertheless, this impacts the fidelity classification of the simulators. Future research should be conducted to address this omission and incorporate the visual environment in fidelity assessments. To support this, researchers need to provide accurate descriptions of the visual environment used, and until this is consistently reported in the literature such evaluations would be limited. Including the visual environment on the fidelity rating scale will provide more accurate ratings of the simulators and could help establish a consistent driving simulator fidelity classification scheme.

Although a meta-analysis was considered, the data drawn from the reviewed literature was not adequate for a meta-analysis. Currently, studies do not use a standard set of measures, or if the same measures are used, one cannot be sure that they have been calculated in the same way. Further, some authors did not report any statistical evidence to support their claims or failed to report sufficient detail to perform a meta-analysis. Even if variables were to be looked at in isolation, currently the number of suitable potential studies would have an insufficient sample size for a rigorous meta-analysis. If the recommendations from this review are adopted, it will be possible to perform a meta-analysis, or series of meta-analyses, in future. This is however a critical matter for consideration and exploring ways of conducting meta-analysis in this area would be worthwhile future research.

Finally, a major consideration for all simulator research is the implication of sampling bias. Although sampling bias is an issue for any research that relies on volunteers, a further issue in simulator studies is simulator sickness. The exact prevalence of simulator sickness is unknown, with reported rates varying from 10% (Kawano et al., 2012; Lee et al. 2003) to 75% (Matas et al., 2015). The presence of simulator sickness is highly influenced by participant factors (e.g., age, gender) and simulator design characteristics such as FOV, screen size, graphic quality, and the presence of a motion base (e.g., Bridgeman et al., 2014; Jager et al., 2014; Lee et al. 2013; Matas et al., 2015; Schreier et al., 2018). This means that samples for driving simulator studies are inherently not representative of all drivers, as one must systematically exclude a sub-group of the population. Researchers tend to be overcautious in the pre-screening of simulator research by eliminating those who have previously experienced symptoms of motion sickness, so as to avoid potentially inducing symptoms in sufferers. This creates a further challenge for the translation and interpretation of research gathered through simulator studies.

4.6. Areas for further research

Currently there are no guidelines for researchers to adhere to when conducting simulator research, and while many adopt similar practices, there is still great variability in the way studies are designed, conducted, and reported. The review highlights various measures that can be taken to ensure comparability in future driving simulator research. First, it is important that standard measures are employed in driving simulator studies. In the current review, no single measure was consistently reported across all items. As they consistently demonstrate absolute validity, it is recommended that speed and/or speed variation be reported in all driving simulator research, as a consistent measure. Speed is selected as it is currently recognised as one of the most reliable variables to determine simulator validity (Bella, 2008; Blaauw, 1982; Mueller, 2015; Reed and Green, 1999), has repeatedly demonstrated absolute validity (see Table 5), is a commonly reported measure across driving research, and would therefore be easy to integrate into most research protocols. This common reference point should also be included as a baseline measure when simulators are being used to measure the impact of a secondary variable on driving performance, such as blood alcohol level. Some items were removed from consideration in the review for not reporting a baseline condition measure of simulator and real-drive (e.g., Engstrom et al., 2005; Filtness et al., 2014).

Second, to allow for further comparisons across different experiments, it is proposed that full statistical results, including descriptive statistics and measures of effect size, be reported. Effect size provides researchers with the opportunity to quantify the impact that driving conditions have on outcome variables. With a reported effect size, researchers can then make face value comparisons between normally distributed groups (Field, 2009). Additionally, researchers should fully report non-significant results; for example, Alm (1996) reported equivalence and differences between the simulator and real-world. This way readers are able to identify potential limitations of the research.

Finally, when describing the simulator used in simulation studies, authors should include detailed information about the simulator's physical, motion, and visual fidelity. Some studies provide this, but in two cases there was no information provided and external information sources had to be reviewed to obtain this information. While the fidelity matrix developed for this review is acknowledged as being an incomplete scale, it offers a guide for the classification. It is also important that simulator descriptions describe key characteristics of the visual driving environment. This should address: the software used, traffic details, the complexity of the visual environment (e.g., city streets versus rural road), if the scenario plays out independent of the driver's actions or if they are determining the course, and – especially in the cases of validation studies – if the route being driven is modelled on a real road that the drivers are familiar with. Where possible comparison images should also be provided, so that readers are able to see the environment driven by participants; if the scenario is modelled off a real area, these images should also be included. By providing this information readers will be able to better appreciate the potential impact of the simulator on the results.

4.7. Conclusion

In summary, although driving simulators allow researchers to examine driving behaviours in safe, and controlled environments, the validity of the results is heavily contingent on the fidelity of the simulator. Although this seems intuitive, importantly, the relationship between simulator fidelity and validity is not straightforward, with some low-fidelity simulators demonstrating acceptable validity on some measures, and some high-fidelity simulators appearing to be invalid on other measures. This, therefore, highlights the need to carefully select appropriate simulator characteristics for the specific research design and aims. In addition, it is concluded that the reporting of driving simulator studies requires improvement, particularly around validation evidence associated with the simulator, the nature of the simulated driving environment, and the outputs of statistical analyses. To ensure the outcomes from such research are indicative of real-world conditions, and to allow for better understanding of the findings, it is envisioned that the suggestions outlined above be adopted.

5. Acknowledgments

We would like to thank the reviewers for their comments and feedback on our manuscript. This work was supported by the Australian Research Council [grant numbers DE150100083 and FT140100681].

6. References

** Indicates items included in the systematic review*

Abe, G., Richardson, J., 2005. The influence of alarm timing on braking response and driver trust in low speed driving. Safety Sci. 43(9), 639–654. https://doi.org/10.1016/j.ssci.2005.04.006.

*Abdel-Aty, M., Yan, X., Radwan, E., Harris, G., Klee, H., 2006. Using the UCF driving simulator as a test bed for high risk locations: Final Report. Center for Advanced Transportation Systems Simulation: University of Central Florida.

Ahlgren, E., Lundqvist, A., Nordlund, A., Aren, C., Rutberg, H., 2003. Neurocognitive impairment and driving performance after coronary artery bypass surgery. Eur. J. Cardiothorac. Surg. 23(3), 334–340. https://doi.org/10.1016/S1010-7940(02)00807-2.

Akinwuntan, A.E., Devos, H., Baker, K., Phillips, K., Kumar, V., Smith, S., Williams, M.J., 2014. Improvement of driving skills in persons with relapsing-remitting multiple sclerosis: a pilot study. Arch. Phys. Med. Rehabil. 95(3), 531–537. https://doi.org/10.1016/j.apmr.2013.08.294.

Allen, R.W., Schwartz, S.H., Hogge, J.R., Stein, A.C., 1978. The effects of alcohol on the driver's decision-making behavior. Volume I. Executive summary and technical report. Report number: DOT HS 803 608. Systems Technology Incorporated.

*Alm, H., 1996. Driving Simulators as Research Tools: A Validation Study Based on the VTI Driving Simulator. Swedish National Road and Transport Research Institute.

Annett, J., 2002. A note on the validity and reliability of ergonomics methods. Theor. Issues Egon. 3(2), 228–232. https://doi.org/10.1080/14639220210124067.

Auberlet, J.-M., Rosey, F., Anceaux, F., Aubin, S., Briand, P., Pacaux, M.-P., Plainchault, P., 2012. The impact of perceptual treatments on driver's behavior: from driving simulator studies to field tests: first results. Accid. Anal. Prev. 45, 91–98. https://doi.org/10.1016/j.aap.2011.11.020.

Ba, Y., Zhang, W., Salvendy, G., Cheng, A.S.K., Ventsislavova, P., 2016. Assessments of risky driving: a Go/No-Go simulator driving task to evaluate risky decision-making and associated behavioral patterns. Appl. Ergon. 52, 265–274. https://doi.org/10.1016/j.apergo.2017.07.020.

*Barker, L., Polson, J., Dupont, P., 1978. Driver screening simulator evaluation program(Final Report). Report number: DOT HS 803 595. Southeastern Oklahoma State University.

*Bedard, M., Parkkari, M., Weaver, B., Riendeau, J., Dahlquist, M., 2010. Assessment of driving performance using a simulator protocol: validity and reproducibility. Am. J. Occup. Ther. 64(2), 336–340. https://doi.org/10.5014/ajot.64.2.336.

Behr, M., Poumarat, G., Serre, T., Arnoux, P.-J., Thollon, L., Brunet, C., 2010. Posture and muscular behaviour in emergency braking: an experimental approach. Accid. Anal. Prev. 42(3), 797–801. https://doi.org/10.1016/j.aap.2009.04.010.

*Bella, F., 2005. Validation of a driving simulator for work zone design. Transport. Res. Rec. 1937, 136–144. https://doi.org/10.3141/1937-19.

*Bella, F., 2008. Driving simulator for speed research on two-lane rural roads. Accid. Anal. Prev. 40(3), 1078–1087. https://doi.org/10.1016/j.aap.2007.10.015.

Bellem, H., Kluver, M., Schrauf, M., Schoner, H.-P., Hecht, H., Krems, J.F., 2017. Can we study autonomous driving comfort in moving-base driving simulators? A validation study. Hum. Factors 59(3), 442–456. https://doi.org/10.1177/0018720816682647.

*Bham, G.H., Leu, M.C., Vallati, M., Mathur, D.R., 2014. Driving simulator validation of driver behavior with limited safe vantage points for data collection in work zones. J. Safety Res. 49, 52–60. https://doi.org/10.1016/j.jsr.2014.02.012.

*Blaauw, G.J., 1982. Driving experience and task demands in simulator and instrumented car: a validation study. Hum. Factors 24(4), 473–486. https://doi.org/10.1177/001872088202400408.

Blana, E., 1996. Driving Simulator Validation Studies: A Literature Review(Working paper 480). Institute of Transport Studies, University of Leeds, Leeds, UK.

Bioulac, S., Franchi, J.A.M., Arnaud, M., Sagaspe, P., Moore, N., Salvo, F., Philip, P., 2017. Risk of motor vehicle accidents related to sleepiness at the wheel: a systematic review and meta-analysis. Sleep 40(10). https://doi.org/10.1093/sleep/zsx134.

Bridgeman, B., Blaesi, S., Campusano, R., 2014. Optical correction reduces simulator sickness in a driving environment. Hum. Factors 56(8), 1472–1481. https://doi.org/10.1177/0018720814533992.

*Branzi, V., Domenichini, L., La Torre, F., 2017. Drivers' speed behaviour in real and simulated urban roads – a validation study. Transp. Res. Part F Traffic Psychol. Behav. 49, 1–17. https://doi.org/10.1016/j.trf.2017.06.001.

Bouchner, P., 2016. Interactive driving simulators: history, design, and their utilization in area of HMI research. Int. J. Syst. Appl. Eng. Dev. 10, 179–188.

Caird, J.K., Horrey, W.J., 2011. Twelve practical and useful questions about driving simulation. In: Fisher, D.L., Rizzo, M., Caird, J.K., Lee, J.D.(Eds.), Handbook of

Driving Simulation for Engineering, Medicine, and Psychology. CRC Press, Boca Raton, FL pp. 5:1-18.

Calhoun, V.D., Pearlson, G.D., 2012. A selective review of simulated driving studies: combining naturalistic and hybrid paradigms, analysis approaches, and future directions. NeuroImage 59(1), 25–35. https://doi.org/10.1016/j.neuroimage.2011.06.037.

*Carter, C.J., Laya, O., 1998. Driver's visual search in a field situation and in a driving simulator. In: Gale, A.G., Brown, I.D., Haslegrave, C.M., Taylor, S.P.(Eds.), Vision in Vehicles—VI. Elsevier Science Ltd, Oxford, England, pp. 21–31.

Charlton, S.G., Starkey, N.J., Perrone, J.A., Isler, R.B., 2014. What’s the risk? A comparison of actual and perceived driving risk. Transp. Res. Part F Traffic Psychol. Behav. 25, 50–64. https://doi.org/10.1016/j.trf.2014.05.003.

Choudhary, P., Velaga, N.R., 2019. Effects of phone use on driving performance: a comparative analysis of young and professional drivers. Safety Sci. 111, 179–187. https://doi.org/10.1016/j.ssci.2018.07.009.

Coeckelbergh, T.R.M., Brouwer, W.H., Cornelissen, F.W., van Wolffelaar, P., Kooijman, A.C., 2002. The effect of visual field defects on driving performance: a driving simulator study. Arch. Opthalmol. 120(11), 1509–1516. https://doi.org/10.1001/archopt.120.11.1509.

Comte, S.L., Jamson, A.H., 2000. Traditional and innovative speed-reducing measures for curves: an investigation of driver behaviour using a driving simulator. Safety Sci. 36(3), 137–150. https://doi.org/10.1016/S0925-7535(00)00037-0.

Cox, D.J., Moore, M., Burket, R., Merkel, R.L., Mikami, A.Y., Kovatchev, B., 2008. Rebound effects with long-acting amphetamine or methylphenidate stimulant medication preparations among adolescent male drivers with attentiondeficit/hyperactivity disorder. J. Child Adolesc. Psychopharmacol. 18(1), 1–10. https://doi.org/10.1089/cap.2006.0141.

*Daurat, A., Sagaspe, P., Motak, L., Taillard, J., Bayssac, L., Huet, N., Authie, C., Mestre, D., Philip, P., 2013. Lorazepam impairs highway driving performance more than heavy alcohol consumption. Accid. Anal. Prev. 60, 31–34. https://doi.org/10.1016/j.aap.2013.08.004.

*Davenne, D., Lericollais, R., Sagaspe, P., Taillard, J., Gauthier, A., Espiec, S., Philip, P., 2012. Reliability of simulator driving tool for evaluation of sleepiness, fatigue and driving performance. Accid. Anal. Prev. 45, 677–682. https://doi.org/10.1016/j.aap.2011.09.046.

de Valck, E., Quanten, S., Cluydts, R., Berckmans, D., 2006. Day versus night driving in real traffic and on a driving simulator during an 800km all-highway drive. Int. J. Vehicle Des. 42(1–2), 119–133. https://doi.org/10.1504/ijvd.2006.010181.

Devos, H., Nieuwboer, A., Tant, M., De Weerdt, W., Vandenberghe, W., 2012. Determinants of fitness to drive in Huntington disease. Neurology 79(19), 1975–1982. https://doi.org/10.1212/WNL.0b013e3182735d11.

Devos, H., Vandenberghe, W., Nieuwboer, A., Tant, M., Baten, G., De Weerdt, W., 2007. Predictors of fitness to drive in people with Parkinson disease. Neurology 69(14), 1434–1441. https://doi.org/10.1212/01.wnl.0000277640.58685.fc.

Devos, H., Vandenberghe, W., Nieuwboer, A., Tant, M., De Weerdt, W., Dawson, J.D., Uc, E.Y., 2013. Validation of a screening battery to predict driving fitness in people with Parkinson's disease. Mov. Disord. 28(5), 671–674. https://doi.org/10.1002/mds.25387.

de Winter, J.C.F., Wieringa, P.A., Dankelman, J., Mulder, M., Van Paassen, M.M., De Groot, S., 2007. Driving simulator fidelity and training effectiveness. In: Proceedings of the

26th European annual conference on human decision making and manual control, Lyngby, Denmark.

Doshi, A., Trivedi, M.M., 2012. Head and eye gaze dynamics during visual attention shifts in complex environments. J. Vision 12(2), 1–16. https://doi.org/10.1167/12.2.9.

*Edwards, D.S., Hahn, C.P., Fleishman, E.A., 1977. Evaluation of laboratory methods for the study of driver behavior: the relation between simulator and street performance. J. Appl. Psychol. 62(5), 559–566. https://doi.org/10.1037/0021-9010.62.5.559.

Engstrom, J., Johansson, E., Ostlund, J., 2005. Effects of visual and cognitive load in real and simulated motorway driving. Transp. Res. Part F Traffic Psychol. Behav. 8(2), 97–120. https://doi.org/10.1016/j.trf.2005.04.012.

Eriksson, A., Banks, V.A., Stanton, N.A., 2017. Transition to manual: comparing simulator with on-road control transitions. Accid. Anal. Prev. 102, 227–234. https://doi.org/10.1016/j.aap.2017.03.011.

Eriksson, A., Stanton, N.A., 2017. Takeover time to highly automated vehicles: noncritical transitions to and from manual control. Hum. Factors 59(4), 689–705. https://doi.org/10.117/0018720816685832.

Field, A., 2009. Discovering Statistics Using IBM SPSS Statistics: and Sex and Drugs and Rock 'N' Roll, third ed. Sage Publications, London.

*Fildes, B.N., Godley, S.T., Triggs, T.J., Jarvis, J., 1997. Perceptual countermeasures: simulator validation study. Report number CR 196(FORS) RR 1/97(RTA). Monash University Accident Research Centre.

Filtness, A.J., Anund, A., Fors, C., Ahlstrom, C., Åkerstedt, T., Kecklund, G., 2014. Sleep-related eye symptoms and their potential for identifying driver sleepiness. J. Sleep Res. 23(5), 568–575. https://doi.org/10.1111/jsr.12163.

*Fors, C., Ahlstrom, C., Anund, A., 2013. Simulator validation with respect to driver sleepiness and subjective experiences: final report of the project SleepEYE II, Part 1. Virtual Prototyping and Assessment by Simulation, VIP: VTI – Swedish National Road and Transport Research Institute.

*Freund, B., Gravenstein, S., Ferris, R., Shaheen, E., 2002. Evaluating driving performance of cognitively impaired and healthy older adults: a pilot study comparing on-road testing and driving simulation. J. Am. Geriatr. Soc. 50(7), 1309–1310. https://doi.org/10.1046/j.1532-5415.2002.50325.x.

Galante, F., Mauriello, F., Montella, A., Pernetti, A., Aria, M., D'Ambrosio, A., 2010. Traffic calming along rural highways crossing small urban communities: driving simulator experiment. Accid. Anal. Prev. 42(6), 1585–1594. https://doi.org/10.1016/j.aap.2010.03.017.

*Galski, T., Bruno, R.L., Ehle, H.T., 1992. Driving after cerebral damage: a model with implications for evaluation. Am. J. Occup. Ther. 46(4), 324–332. https://doi.org/10.5014/ajot.46.4.324.

Gastaldi, M., Rossi, R., 2011. A methodology for calibrating road link travel time functions using data from driving simulator experiments. Proc. Soc. Behav. Sci. 20, 656–665. https://doi.org/10.1016/j.sbspro.2011.08.073.

Goode, N., Salmon, P.M., Lenne, M.G., 2013. Simulation-based driver and vehicle crew training: applications, efficacy and future directions. Appl. Ergon. 44, 435–444. https://doi.org/10.1016/j.apergo.2012.10.007.

Greenberg, J., Blommer, M., 2012. Physical fidelity of driving simulators. In: Fisher, D.L., Rizzo, M., Caird, J.K., Lee, J.D.(Eds.), Handbook of Driving Simulation for Engineering, Medicine, and Psychology. CRC Press, Boca Raton, FL pp. 7:1-24.

Hakamies-Blomqvist, L., Heriksson, P., Lundberg, C., Ostlund, J., 2001. Strategical compensation and driving behavior in older drivers. VTI Report 470. Swedish National Road and Transport Research Institute.

*Hallvig, D., Anund, A., Fors, C., Kecklund, G., Karlsson, J.G., Wahde, M., Åkerstedt, T., 2013. Sleepy driving on the real road and in the simulator: a comparison. Accid. Anal. Prev. 50, 44–50. https://doi.org/10.1016/j.aap.2012.09.033.

*Harms, L., 1996. Driving performance on a real road and in driving simulator: results of a validation study. In: Gale, A.G., Brown, I.D., Haslegrave, C.M., Taylor, S.P.(Eds.), Vision in Vehicles – V. Elsevier, Amsterdam, pp. 19–26.

Havlikova, M., Sediva, S., 2012. Monitoring of driver's fatigue: statistical analysis of test driver trajectories in various environments and in different driver stress degrees. IFAC Proc. Vol. 45(7), 146–150. https://doi.org/10.3182/20120523-3-CZ-3015.00030.

Helland, A., Jenssen, G.D., Lervag, L.-E., Moen, T., Engen, T., Lydersen, S., Slordal, L., 2016. Evaluation of measures of impairment in real and simulated driving: results from a randomized, placebo-controlled study. Traffic Inj. Prev. 17(3), 245–250. https://doi.org/10.1080/15389588.2015.1065975.

*Helland, A., Jenssen, G.D., Lervag, L.-E., Westin, A.A., Moen, T., Sakshaug, K., Slordal, L., 2013. Comparison of driving simulator performance with real driving after alcohol intake: a randomised, single blind, placebo-controlled, cross-over trial. Accid. Anal. Prev. 52, 9–16. https://doi.org/10.1016/j.aap.2012.12.042.

Helman, S., Reed, N., 2015. Validation of the Driver Behaviour Questionnaire using behavioural data from an instrumented vehicle and high-fidelity driving simulator. Accid. Anal. Prev. 75, 245–251. https://doi.org/10.1016/j.aap.2014.12.008.

Hird, M.A., Egeto, P., Fischer, C.E., Naglie, G., Schweizer, T.A., 2016. A systematic review and meta-analysis of on-road simulator and cognitive driving assessment in

Alzheimer’s disease and mild cognitive impairment. J. Alzheimers Dis. 53(2), 713–729. https://doi.org/10.3233/JAD-160276.

Hogan, H.A., Szeto, A.Y., 1982. A simulator for objectively evaluating prospective drivers of the Scott van. Bull. Prosthet. Res. 10, 19–27.

*Hou, Y., Zhao, Y., Hulme, K.F., Huang, S., Yang, Y., Sadek, A.W., Qiao, C., 2014. An integrated traffic-driving simulation framework: design, implementation, and validation. Trans. Res. Part C Emerg. Technol. 45, 138–153. https://doi.org/10.1016/j.trc.2014.02.004.

*Hulme, K.F., Thorpe, L., 2013. A simulation-based assessment approach to increase safety among senior drivers.(Project no. 49111-12-23). New York State Center for Engineering Design and Industrial Innovation(NYSCEDII): University at Buffalo.

Inman, V.W., Davis, G.W., El-Shawarby, I., Rakha, H.A., 2008. Test track and driving simulator evaluations of warnings to prevent right-angle crashes at signalized intersections(Report no. FHWA-HRT-08/070). FHWA Office of Safety R&D, Human Centered Systems Team: Federal Highway Administration.

Jager, M., Gruber, N., Muri, R., Mosimann, U.P., Nef, T., 2014. Manipulations to reduce simulator-related transient adverse health effects during simulated driving. Med. Biol. Eng. Comput. 52(7), 601–610. https://doi.org/10.1007/s11517-014-1162-x.

Jeihani, M., NarooieNezhad, S., Kelarestaghi, K.B., 2017. Integration of a driving simulator and a traffic simulator case study: exploring drivers' behavior in response to variable message signs. IATSS Res. 41(4), 164–171. https://doi.org/10.1016/j.iatsr.2017.03.001.

*Johnson, M.J., Chahal, T., Stinchcombe, A., Mullen, N., Weaver, B., Bedard, M., 2011. Physiological responses to simulated and on-road driving. Int. J. Psychophysiol. 81(3), 203–208. https://doi.org/10.1016/j.ijpsycho.2011.06.012.

Kaplan, R.M., Saccuzzo, D.P., 2009. Psychological Testing: Principles. Applications, and Issues, seventh ed. Wadsworth, Belmont, CA.

Kaptein, N.A., Theeuwes, J., Van Der Horst, R., 1996. Driving simulator validity: some considerations. Transport. Res. Rec. 1550, 30–36.

Kawano, N., Iwamoto, K., Ebe, K., Aleksic, B., Noda, A., Umegaki, H., Ozaki, N., 2012. Slower adaptation to driving simulator and simulator sickness in older adults. Aging Clin. Exp. Res. 24(3), 285–289. https://doi.org/10.1007/BF03325260.

*Klee, H., Bauer, C., Radwan, E., Al-Deek, H., 1999. Preliminary validation of driving simulator based on forward speed. Transport. Res. Rec. 1689, 33–39. https://doi.org/10.3141/1689-05.

Lauer, A.R., 1960. The Psychology of Driving: Factors of Traffic Enforcement. Springfield: C. C. Thomas.

Lauer, A.R., Suhr, V.W., 1958. Development of a criterion for driving performance. In: Driver characteristic and behavior studies. HRB Bull. No. 172, NAS-NRC Publ. No. 532. National Academy of Sciences, National Research Council, Oxford, England, pp. 1–8.

Lauridsen, M.M., Thacker, L.R., White, M.B., Unser, A., Sterling, R.K., Stravitz, R.T., Bajaj, J.S., 2016. In patients with cirrhosis, driving simulator performance is associated with real-life driving. Clin. Gastroenterol. Hepatol. 14, 747–752. https://doi.org/10.1016/j.cgh.2015.11.007.

*Lee, H.C., Cameron, D., Lee, A.H., 2003. Assessing the driving performance of older adult drivers: on-road versus simulated driving. Accid. Anal. Prev. 35(5), 797–803. https://doi.org/10.1016/S0001-4575(02)00083-0.

*Lee, J.D., Hoffman, J.D., Brown, T.L., McGehee, D.V., 2002. Comparison of driver braking responses in a high-fidelity driving simulator and on a test track(Report no. DOT HS 809 447). Human Factors and Vehicle Safety Research Program: University of Iowa.

*Lee, J.D., Ward, N., Boer, E., Brown, T.L., Balk, S.A., Ahmad, O., 2013. Exploratory Advanced Research: Making Driving Simulators More Useful for Behavioral Research: Simulator Characteristics Comparison and Model-based Transformation. National Advanced Driving Simulator: University of Iowa.

Lew, H.L., Poole, J.H., Lee, E.H., Jaffe, D.L., Huang, H.-C., Brodd, E., 2005. Predictive validity of driving-simulator assessments following traumatic brain injury: a preliminary study. Brain Inj. 19(3), 177–188. https://doi.org/10.1080/02699050400017171.

*Li, J., Zhao, X., Xu, S., Ma, J., Rong, J., 2013. The study of driving simulator validation for physiological signal measures. Proc. Soc. Behav. Sci. 96, 2572–2583. https://doi.org/10.1016/j.sbspro.2013.08.288.

Lin, T.W., Hwang, S.L., Green, P.A., 2009. Effects of time-gap settings of adaptive cruise control(ACC) on driving performance and subjective acceptance in a bus driving simulator. Safety Sci. 47(5), 620–625. https://doi.org/10.1016/j.ssci.2008.08.004.

Lundqvist, A., Gerdle, B., Ronnberg, J., 2000. Neuropsychological aspects of driving after a stroke: in the simulator and on the road. Appl. Cogn. Psychol. 14(2), 135–150. https://doi.org/10.1002/(SICI)1099-0720(200003/04)14:2<135::AID-ACP628>3.0.CO;2-S.

Marcotte, T.D., Wolfson, T., Rosenthal, T.J., Heaton, R.K., Gonzalez, R., Ellis, R.J., Grant, I., 2004. A multimodal assessment of driving performance in HIV infection. Neurology 63(8), 1417–1422. https://doi.org/10.1212/01.WNL.0000141920.33580.5D.

Matas, N.A., Nettelbeck, T., Burns, N.R., 2015. Dropout during a driving simulator study: a survival analysis. J. Safety Res. 55, 159–169. https://doi.org/10.1016/j.jsr.2015.08.004.

*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. J. Safety Res. 42(4), 267–275. https://doi.org/10.1016/j.jsr.2011.06.004.

*McAvoy, D.S., Schattler, K.L., Datta, T.K., 2007. Driving simulator validation for night-time construction work zone devices. Transport. Res. Rec. 2015, 55–63. https://doi.org/10.3141/2015-07.

Mecheri, S., Lobjois, R., 2018. Steering control in a low-cost driving simulator: a case for the role of virtual vehicle cab. Hum. Factors 60(5), 719–734. https://doi.org/10.1177/0018720818769253.

Merat, N., Anttila, V., Luoma, J., 2005. Comparing the driving performance of average and older drivers: the effect of surrogate in-vehicle information systems. Transp. Res. Part F Traffic. Psychol. Behav. 8(2), 147–166. https://doi.org/10.1016/j.trf.2005.04.011.

*Meuleners, L., Fraser, M., 2015. A validation study of driving errors using a driving simulator. Transp. Res. Part F Traffic. Psychol. Behav. 29, 14–21. https://doi.org/10.1016/j.trf.2014.11.009.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., Prisma Group, 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoSMed. 6(7). https://doi.org/10.1371/journal.pmed.1000097.

*Mueller, J.A., 2015. Driving in a Simulator Versus On-Road: The Effect of Increased Mental Effort While Driving on Real Roads and a Driving Simulator. PhD dissertation. Montanta State University.

Mullen, N., Charlton, J.L., Devlin, A., Bedard, M., 2012. Simulator validity: behaviors observed on the simulator and on the road. In: Fisher, D.L., Rizzo, M., Caird, J.K., Lee, J.D.(Eds.), Handbook of Driving Simulation for Engineering, Medicine, and Psychology. CRC Press, Boca Raton, FL pp. 13:1-18.

Neisser, U., 1976. Cognition and reality: principles and implications of cognitive psychology. MacMillan Publishers, New York.

O'Hern, S., Oxley, J., Stevenson, M., 2017. Validation of a bicycle simulator for road safety research. Accid. Anal. Prev. 100, 53–58. https://doi.org/10.1016/j.aap.2017.01.002.

*Philip, P., Sagaspe, P., Taillard, J., Valtat, C., Moore, N., Åkerstedt, T., Charles, A., Bioulac, B., 2005. Fatigue, sleepiness, and performance in simulated versus real driving conditions. Sleep 28(12), 1511–1516. https://doi.org/10.1093/sleep/28.12.1511.

Pietrucha, M.T., Hostetter, R.S., Staplin, L., Obermeyer, M., 1996. Pavement markings and delineation for older drivers. Volume I: Final report. McLean, Virginia: Turner-Fairbank Highway Research Center, U. S. Department of Transportation.

Polders, E., Cornu, J., De Ceunynck, T., Daniels, S., Brijs, K., Brijs, T., Hermans, E., Wets, G., 2015. Drivers' behavioural responses to combined speed and red-light cameras. Accid. Anal. Prev. 81, 153–166. https://doi.org/10.1016/j.aap.2015.05.006.

Radwan, E., Al-Deek, H., Bauer, C., Klee, H., 1999. Driving simulator technologies(Final report). University of Central Florida. Ranney, T.A., 2012. Psychological fidelity: perception of risk. In: Fisher, D.L., Rizzo, M., Caird, J.K., Lee, J.D.(Eds.), Handbook of Driving Simulation for Engineering, Medicine, and Psychology. CRC Press, Boca Raton, FL pp. 9:1-13.

*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. https://doi.org/10.1080/001401399185117.

*Risto, M., Martens, M.H., 2014. Driver headway choice: a comparison between driving simulator and real-road driving. Transp. Res. Part F Traffic. Psychol. Behav. 25, 1–9. https://doi.org/10.1016/j.trf.2014.05.001.

*Santos, J., Merat, N., Mouta, S., Brookhuis, K., de Waard, D., 2005. The interaction between driving and in-vehicle information systems: comparison of results from laboratory, simulator and real-world studies. Transp. Res. Part F Traffic. Psychol. Behav. 8(2), 135–146. https://doi.org/10.1016/j.trf.2005.04.001.

Schreier, D.R., Banks, C., Mathis, J., 2018. Driving simulators in the clinical assessment of fitness to drive in sleepy individuals: a systematic review. Sleep Med. Rev. 38, 86–100. https://doi.org/10.1016/j.smrv.2017.04.004.

*Senserrick, T.M., Brown, T., Quistberg, D.A., Marshall, D., Winston, F.K., 2007. Validation of simulated assessment of teen driver speed management on rural roads. Annu. Proc. Assoc. Adv. Automot. Med. 51, 525–536.

*Shechtman, O., Classen, S., Awadzi, K., Mann, W., 2009. Comparison of driving errors between on-the-road and simulated driving assessment: a validation study. Traffic Inj. Prev. 10(4), 379–385. https://doi.org/10.1080/15389580902894989.

Stanton, N.A., 2016. On the reliability and validity of, and training in, ergonomics methods: a challenge revisited. Theor. Issues Egon. 17(4), 345–353. https://doi.org/10.1080/1463922X.2015.11176882016.

Stanton, N.A., Young, M.S., Walker, G.H., Turner, H., Randle, S., 2001. Automating the driver's control tasks. Int. J. Cogn. Ergon. 5(3), 221–236. https://doi.org/10.1207/S15327566IJCE0503_5.

Strayer, D.L., Turrill, J., Cooper, J.M., Coleman, J.R., Medeiros-Ward, N., Biondi, F., 2015. Assessing cognitive distraction in the automobile. Hum. Factors 57, 1300–1324. https://doi.org/10.1177/001872081575149.

Suhr, V., 1957. Comparison of laboratory and field studies in the estimation of driving speed. Proc. Iowa Acad. Sci. 64, 546–552.

Szlyk, J.P., Alexander, K.R., Severing, K., Fishman, G.A., 1992. Assessment of driving performance in patients with retinitis pigmentosa. Arch. Opthalmol. 110(12), 1709–1713. https://doi.org/10.1001/archopht.1992.01080240049027.

Szlyk, J.P., Pizzimenti, C.E., Fishman, G.A., Kelsch, R., Wetzel, L.C., Kagan, S., Ho, K., 1995. A comparison of driving in older subjects with and without age-related macular degeneration. Arch. Opthalmol. 113(8), 1033–1040. https://doi.org/10.1001/archopht.1995.01100080085033.

*Tornros, J., 1998. Driving behaviour in a real and a simulated road tunnel—a validation study. Accid. Anal. Prev. 30(4), 497–503. https://doi.org/10.1016/S0001-4575(97)00099-7.

van Erp, J.B.F., Padmos, P., 2003. Image parameters for driving with indirect viewing systems. Ergonomics 46(15), 1471–1499. https://doi.org/10.1080/0014013032000121624.

van Wolffelaar, P., van Zomeren, E., Brouwer, W., Rothengatter, T., 1988. Assessment of fitness to drive of brain damaged persons. Papers Presented at the 2nd International Conference on Road Safety.

*Veldstra, J.L., Bosker, W.M., de Waard, D., Ramaekers, J.G., Brookhuis, K.A., 2015. Comparing treatment effects of oral THC on simulated and on-the-road driving performance: testing the validity of driving simulator drug research. Psychopharmacology(Berl.) 232(16), 2911–2919. https://doi.org/10.1007/s00213-015-3927-9.

Volkerts, E.R., Van Laar, M.W., Van Willigenburg, A.P., Plomp, T.A., Maes, R.A.A., 1992. A comparative study of on-the-road and simulated driving performance after nocturnal treatment with lormetazepam 1 mg and oxazepam 50 mg. Hum. Psychopharm. Clin. Exp. 7(5), 297–309. https://doi.org/10.1002/hup.470070502.

*Wade, M.G., Hammond, C., 1998. Simulation validation: evaluating driver performance in simulation and the real world. Report no. MN/RC-1998-28. University of Minnesota, Hum Factors Research Laboratory.

Wang, J., Li, K., Lu, X.-Y., 2014. Effect of Hum Factors on driver behavior. In: Chen, Y., Li, L.(Eds.), Advances in Intelligent Vehicles. Academic Press, Boston, pp. 111–157.

*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. https://doi.org/10.1080/00140130903464358.

Wilson, G.F., 2002. An analysis of mental workload in pilots during flight using multiple psychophysiological measures. Int. J. Aviat. Psychol. 12(1), 3–18. https://doi.org/10.1207/S15327108IJAP1201_2.

Yan, X., Abdel-Aty, M., Radwan, E., Wang, X., Chilakapati, P., 2008. Validating and driving simulator using surrogate safety measures. Accid. Anal. Prev. 40, 274–288.https://doi.org/10.1016/j.aap2007.06.007.

*Yun, M., Zhao, J., Zhao, J., Weng, X., Yang, X., 2017. Impact of in-vehicle navigation information on lane-change behavior in urban expressway diverge segments. Accid. Anal. Prev. 106, 53–66. https://doi.org/10.1016/j.aap.2017.05.025.

Zhang, H., Chavarriaga, R., Khaliliardali, Z., Gheorghe, L., Iturrate, I., Millan, J., 2015. EEG-based decoding of error-related brain activity in a real-world driving task. J. Neural Eng. 12(6), 066028. https://doi.org/10.1088/1741-2560/12/6/066028.

*Zoller, I., Abendroth, B., Bruder, R., 2019. Driver behaviour validity in driving simulators – analysis of the moment of initiation of braking at urban intersections. Transp. Res. Part F: Traffic Psychol. Behav. 61, 120–130. https://doi.org/10.1016/j.trf.2017.09.008.

Table 1. Initial search results for systematic literature review.

Database	Number of Search Results for Each Search Term				
	driv* OR vehicle OR automobile OR car OR truck OR lorry OR van OR motorcycl* (1)	simulat* (2)	valid* OR evaluat* OR compar* (3)	1 + 2	(1 + 2) + 3
PsycInfo	233,581	51,006	540,112	4,289	1,852
PubMed	422,228	377,776	6,184,543	21,015	8,547
ScienceDirect	103,086	191,378	579,078	18,319	8,807
TRID	>15,000	>15,000	>15,000	>15,000	3,909

Note. Database search performed 8–10 August 2016.

Table 2. Eligibility criteria for search results.

Inclusion criteria

- article must contain:
  1. simulated motor vehicle driving in a vehicle (i.e., not truck, motorbike, or bicycle)
  2. measure of real-world driving (i.e., on-road drive, self-reported driving behaviour, neuropsychological assessment, accident or driving history etc.)
- originally published in English
- full article (i.e., not published abstract only)

Exclusion criteria

- not related to driving (i.e., mathematical simulations, physiology, medicine etc.)
- no original data (e.g., meta-analysis, editorial etc)
- driving simulation and real driving not compared
- full text not available

Additional inclusion criteria applied in review

- simulated driving in a car (i.e., not truck, motorbike, or bicycle)

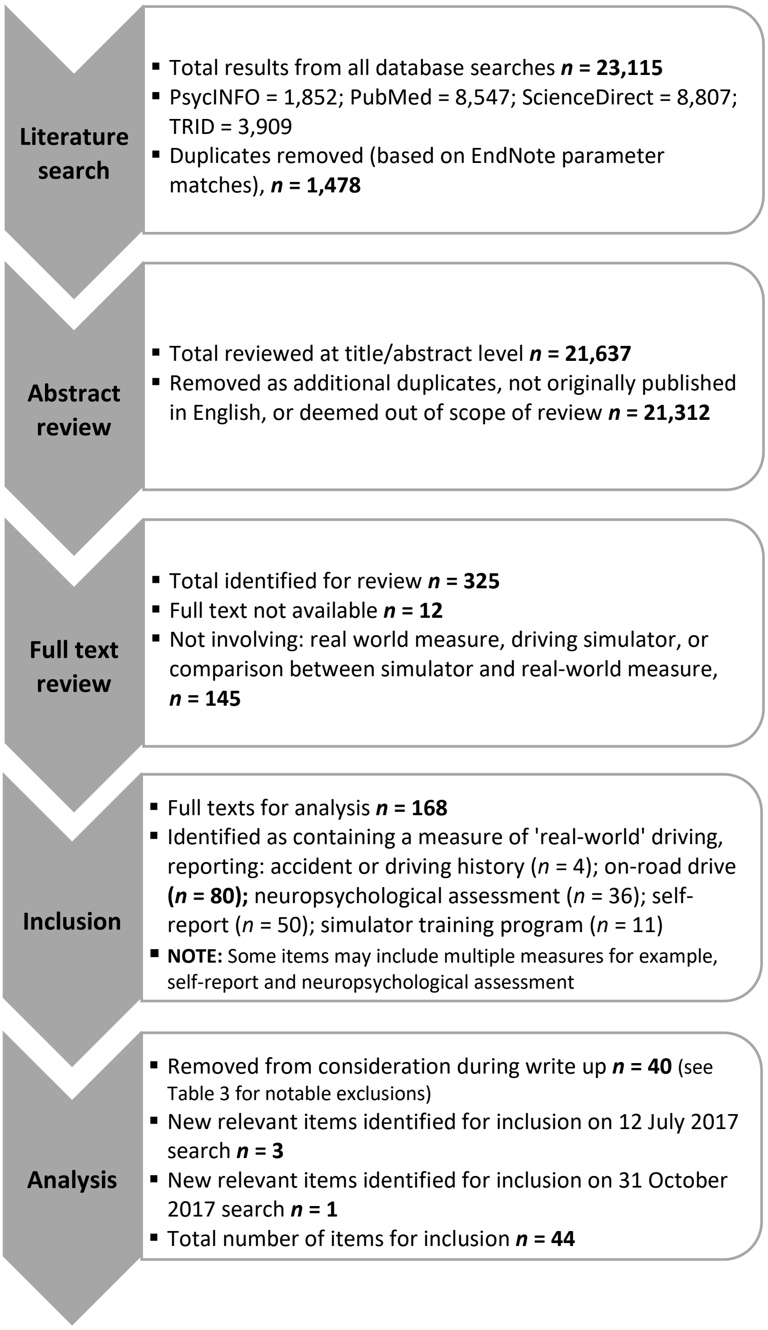

Figure 1. Systematic process of literature review.

Table 3. Notable exclusions from the systematic review.

Reason for exclusion	Item reference
Automated driving simulation	Bellem et al. (2017); Eriksson, Banks, & Stanton (2017)
Duplication of another item or not empirical data	Helland et al. (2016); Helman & Reed (2015); Radwan et al. (1999)
Does not report comparison between simulator and real drive on driving performance measure in baseline (control) condition	Ahlgren et al. (2003); Akinwuntan et al. (2014); Auberlet et al. (2012); Cox et al. (2008); de Valck et al. (2006); Devos et al. (2013); Doshi & Trivedi (2012); Engström, Johansson, & Östlund (2005); Galante et al. (2010); Lauer & Suhr (1958); Lew et al. (2005); Lundquist et al. (2000); Marcotte et al. (2004); Merat et al. (2005); Pietrucha et al. (1996); Szlyk et al. (1995); Wang et al. (2014); Zhang et al. (2015)
Focus on secondary measure or task; not driving performance measure	Allen et al. (1978); Behr et al. (2010); Charlton et al. (2014); Filtness et al. (2014); Gastaldi & Rossi (2011); Havlikova & Sediva (2012); Inman et al. (2008); Stanton et al. (2001); Strayer et al. (2015); van Erp & Padmos (2003);
Not in English	Hakamies-Blomqvist et al. (2001)
Simulator composition is not comparable with others in the review	Hogan & Szeto (1982); O'Hern et al. (2017); Suhr (1957); Volkerts et al. (1992)
Simulator used as predictor of fitness to drive	Coeckelbergh et al. (2002); Devos et al. (2007); Devos et al. (2012)
Systematic reviews including but not comparing simulator and on-road studies	Bioulac et al. (2017); Hird et al. (2016)

Table 4. Rating system used for simulator fidelity.

	Score				
Fidelity measure	1	2	3	4	5
Visual	Single PC screen	Single projector or PC screen >25 inches	<180° FOV with multiple screens	180-270° FOV; multiple PC screens	> 270° FOV; projector screens
Motion	No motion base		Lower degrees of motion (<6 or partial plate)		Motion-base on Full motion plate
Physical	Computer based-simulation using keyboard or joystick	Computer-based simulation with steering wheel	"Arcade" vehicle – car seat with steering wheel and pedals	Vehicular controls; no or incomplete cab	Full vehicular cab and controls

Table 5 Studies Reporting Speed Measures when Comparing Simulator (SIM) and On-Road (OR) Drives.

Study	Fidelity	Measure	Contrast	Results	Validity
Abdel-Aty et al. (2006)	HIGH	Mean speed	Between	No significant differences for comparisons in any road condition.	Absolute, all conditions
Bella (2005)	MED	Mean speed	Between	Non-significant differences between SIM and OR for all measurement sites.	Absolute, all locations
Bella (2008)	MED	Mean speed	Between	For 9 of 11 measured sites, speed in the SIM was not significantly different to OR speeds.	Absolute, most locations
Bham et al. (2014)	MED	Mean speed	Between	No significant differences between speeds in SIM and OR.	Absolute, all conditions
Klee et al. (1999)	MED	Mean speed	Within	No significant differences between SIM and real drive at 10 of 16 speed locations; *p*-values and effect sizes not provided.	Absolute, some locations
Blaauw (1982)	LOW	Mean speed Speed variation	Within	Consistent between sim and OR; results not reported. Inexperienced drivers showed higher variation than experienced across both	Relative
Fildes et al. (1997)	HIGH	Speed variation	Between	Significant correlations between SIM and OR for approach speed changes between control and treatment at 3 of 4 test sites: stop sign, $r = .40$; left-curve, $r = .48$; right-curve, $r = .52$; not significant at roundabout, $r = -.18$.	Relative, some conditions
Lee et al. (2013)	HIGH LOW	Mean speed	Between	All SIMs showed high R^2 values (NADS: .88; WTI: .86; FHWA: .78; miniSim: .83), indicating strong relationship between SIM and OR	Relative
Reed & Green (1999)	MED	Mean speed	Within	Comparable speeds between SIM and OR. Small correlation found for SD of speed ($r = .19$). No significance values provided.	Relative
Branzi et al. (2017)	HIGH	Mean speed Speed variation	Between	Using combination of parametric and non-parametric tests; not significantly different for 13 of 21 sites. Increases and decreases in speed followed similar trends in each section across the SIM and OR.	Mixed (speed: absolute some locations; speed variation: relative).
Harms (1996)	MED	Mean speed	Within	Despite speed appearing generally higher in SIM than OR there was no significant difference, $F(1,36) = 3.67$, $p > .060$. Strong correlations between conditions: Sim 1–Field 2: $r = .88$; Sim 2–Field 1: $r = .86$.	Mixed (absolute and relative)
Alm (1996)	MED	Mean speed Speed variation	Within	No significant differences in mean speed; results not reported. Speed variation significantly larger in SIM than OR, $F(2,48) = 10.24$, $p = .0002$.	Mixed (speed: absolute; speed variation: none).

Study	Fidelity	Measure	Contrast	Results	Validity
McAvoy et al. (2007)	MED	Mean speed	Between	Non-significant difference between SIM and OR speeds ($p > .050$) for middle zone of control site and beginning and end zones of test site. Significant difference between SIM and field speeds for beginning and end zones of control sites and middle of test site. Significant differences between the expected and observed speeds in beginning and middle of test site.	Mixed (absolute, some locations; none, some locations)
Carter & Laya (1998)	LOW	Max and min speed	Within	SIM had significantly lower minimum speeds, $F(1,48) = 15.21$, $p < .001$, and significantly higher maximum speeds, $F(1,48) = 47.85$, $p < .001$.	None
Fors et al. (2013)	HIGH	Mean speed	Within	Significantly slower OR than SIM. Also significant time * drive type interaction; speed differences between OR and SIM larger at night.	None
Hallvig et al. (2013)	HIGH	Mean speed	Within	Significantly higher speeds in SIM than OR, $F(1,9) = 24.7$, $p < .001$.	None
Santos et al. (2005)	LOW MED	Mean speed Speed variation	Between	Significant differences in mean speed between LEEDS SIM and OR, $t(46) = 5.12$, $p < .010$. Significantly less speed variation in low-fidelity lab SIM than OR, $t(46) = 15.11$, $p < .010$.	None
Senserrick et al. (2007)	HIGH	Mean speed Speed variation	Within	Significant main effect for driving location, $F(1, 16) = 13.53$, $p < .001$. Drivers varied speed significantly more OR than SIM.	None
Törnros (1998)	MED	Mean speed	Within	Higher speed in SIM than real tunnel, $p < .001$, $\omega^2 = .36$.	None
Wang et al. (2010)	MED	Mean speed	Between	Significantly faster in SIM than in field, $F(1,55) = 74.00$, $p < .001$.	None
Zöller et al. (2019)	LOW MED	Mean speed	Within	For all SIM variations, drivers drove at a higher speed in the SIM than in the field (*p*-values not reported).	None

Table 6. Studies reporting lateral position measures between simulator (SIM) and on-road (OR) drives.

Study	Fidelity	Measure	Contrast	Results	Validity
Veldestra et al. (2015)	MED	Lane pos. variation	Within	No significant difference overall between SIM and OR, $F(1,19) = 0.15$, $p = .710$, $\eta_p^2 = .008$.	Absolute
Reed & Green (1999)	MED	Mean lane pos., mean lateral speed, SD lane pos.	Within	Moderate to large correlations between OR drive and SIM for lane position ($r = .43$); mean lateral speed ($r = .76$) and SD of lane position ($r = .59$).	Relative
Alm (1996)	MED	Lateral pos. Lateral pos. variation	Within	No significant difference between SIM and OR for mean lateral position; results not reported. Significantly more lateral variation in static SIM vs OR or moving base SIM, $F(2,48) = 9.12$, $p = .0004$.	Mixed (lateral position: absolute; lateral position variation: none).
Harms (1996)	MED	Lateral pos. Distance to centreline	Within	Moderate correlations for lateral position; a less consistent driving pattern between driving conditions: Sim 1–Field 2: $r = .51$; Sim 2–Field 1: $r = .47$. Significantly closer to centreline in SIM than OR, $F(1,32) = 741.44$, $p < .001$.	Mixed (lateral position: relative; distance to centreline: none)
Blaauw (1982)	LOW	Lateral pos.	Within	Mean and SD significantly larger in the sim than in car, $p < .010$.	None
Daurat et al. (2013)	MED	Lane pos. variation	Within	SD of lane position significantly higher in SIM than real drive, $F(1,13) = 24.90$, $p < .001$, $\eta_p^2 = .65$.	None
Hallvig et al. (2013)	HIGH	Lateral pos.	Within	Significant differences between SIM and OR for lateral pos. mean, $F(1,9) = 7.6$; $p < .050$, and SD, $F(1,9) = 18.2$, $p < .010$.	None
Helland et al. (2013)	HIGH	Lateral pos. variation	Mixed	Greater in SIM (29.4 cm) than OR (22.3 cm). Significant difference in SIM (but not OR) for curved sections than straight sections	None
Törnros (1998)	MED	Mean lateral pos.	Within	Drivers positioned themselves further from the tunnel wall OR than in the SIM both when road was straight and curved.	None
Wade & Hammond (1998)	LOW	Mean lateral pos.	Within	Significantly higher in SIM than OR, significance value was absent	None
Wang et al. (2010)	MED	Lateral pos. variation	Between	SIM was significantly higher than OR, $F(1,47) = 4.80$, $p = .033$.	None

Table 7. Studies reporting line crossing and lane change behaviours simulator (SIM) and on-road (OR) drives.

Study	Fidelity	Measure	Contrast	Results	Validity
Fors et al. (2013)	HIGH	Line crossings	Within	ANOVA showed no significant main effects of drive type; results not presented	Absolute
Daurat et al. (2013)	MED	Line crossings	Within	No significant difference between OR and SIM in placebo condition ($p > .07$)	Absolute
Davenne et al. (2012)	LOW	Line crossings	Between	Significantly more line crossings in SIM than OR, $F(1,132) = 27.72$, $p < .001$.	None
Philip et al. (2015)	LOW	Line crossings	Within	Significantly more line crossings in SIM than OR, $F(1,10) = 60.01$; $p < .001$.	None
Yun et al. (2017)	HIGH	Lane change behaviour	Between	Nonparametric tests found no significant differences for merging gap and lane change position between OR and SIM; results not presented	Absolute

Table 8. Studies reporting driving errors or an overall driving score (as rated by an external evaluator or simulator program).

Study	Fidelity	Measure	Contrast	Results	Validity
Barker et al. (1978)	LOW	Overall mean score	Within	No significant difference between SIM scores and Oklahoma Driver Licence Test (OR); $t(45) = 0.56$, $p > .01$.	Absolute
Meuleners & Fraser (2015)	MED	Total errors	Within	No significant difference in error scores (results not reported).	Absolute
Bédard et al. (2010)	LOW	Demerit points	Within	Significant correlation between demerit points awarded OR and SIM, $r = .74$.	Relative
Freund et al. (2002)	LOW	Sim errors and OR performance	Within	Significant negative correlations between OR performance and overall ($r = -.67$), hazardous ($r = -.83$) and lethal ($r = -.82$) SIM errors. **Fewer SIM errors = higher OR driving performance.*	Relative
Galski et al. (1992)	LOW	Sim errors and OR performance	Within	Significant correlation between signalling errors in SIM and evaluations made of OR drive, $r = -.64$.	Relative
Lee et al. (2003)	LOW	Overall performance	Between	SIM and OR performance positively correlated, $r = .72$.	Relative
Mayhew et al. (2011)	LOW	Total errors	Within	Significant correlations between errors recorded by examiners OR and in SIM, tau = .23; $p = .010$.	Relative
Edwards et al. (1977)	MED	Overall score	Within	No significant correlations for any component between OR and either SIM (no statistics reported).	None
Hulme & Thorpe (2013)	MED	Total errors	Within	More raw errors for SIM than OR; no statistical comparison reported.	None
Schectman et al. (2009)	MED	Total errors	Within	Significantly more errors OR than SIM, $F(1,77) = 7.40$, $p < .010$.	None

Table 9. Studies reporting other measures of driving or physical performance, comparing simulator and on-road (OR) drives.

Study	Fidelity	Measure	Contrast	Results	Validity
				Driving Performance Measures	
Hou et al. (2014)	HIGH	Travel time	Within	No mean differences between SIM and field test.	Absolute
Johnson et al. (2011)	MED	Travel time	Within	No significant differences between SIM and OR, $t(8) = 0.58$, $p = .578$.	Absolute
Risto & Martens (2014)	MED	Vehicle headway	Within	No significant differences between SIM and OR; no p-values presented	Absolute
Lee et al. (2002)	HIGH	Braking behaviour	Between	"Quite similar" braking profiles appear between SIM and OR; no statistical comparisons reported.	Relative
Zöller et al. (2019)	LOW MED	Braking behaviour	Within	Drivers initiated braking in anticipation of a turn earlier OR than in SIM (all types); all 20 comparisons significant.	None
Reed & Green (1999)	MED	Steering	Within	SD steering wheel position correlated well between car and SIM, $r = .74$.	Relative
Mueller (2015)	HIGH	Steering	Within	SD steering angle and steering reversal frequency were significantly higher in SIM than OR (post hoc tests).	None
				Non-Driving Measures	
Li et al. (2013)	MED	EEG	Between	No significant difference found at 7 of 8 locations between SIM and OR; "good correspondence" of beta waves.	Absolute
Li et al. (2013)	MED	Heart rate	Between	Not significantly different between SIM and OR at 7 of 8 locations; results not reported.	Absolute
Johnson et al. (2011)	MED	Heart rate	Within	MaxHR significantly higher during OR than SIM, $F(1,8) = 9.66$, $p = .015$. Non-significant correlation between drives for mean HR, $r = .49$, $p = .177$.	None
Alm (1996)	MED	Mental workload	Within	SIM rated higher for physical demands, $F(1,34) = 4.83$, $p = .035$, effort, $F(1,34) = 10.06$, $p = .003$, and frustration, $F(1,34) = 6.82$, $p = .013$.	None
Mueller (2015)	HIGH	Mental workload	Within	Simulated driving had lower self-reported performance and higher mental effort than OR; results not reported	None
Carter & Laya (1998)	LOW	Eye fixations	Within	In SIM drivers had: more fixations per second, $F(1,48) = 4.93$, $p < .040$, fewer fixation zones, $F(1,48) = 9.26$, $p < .010$, more time overall in fixation, $F(1,48) = 10.85$, $p < .010$, and more time fixating on dashboard, $F(1,48) = 10.29$, $p < .010$.	None
Fors et al. (2013)	HIGH	Eye fixations	Within	Drivers spent more time looking ahead on the road in OR than SIM.	None
Mueller (2015)	HIGH	Eye fixations	Within	Horizontal and vertical gaze dispersion was significantly greater in the SIM than OR. More on-road and less off-road fixations per minute OR than SIM.	None

Supplementary Materials: Systematic Review of Simulator Validation Studies

List of Included Items, Study Details and Measures.

Reference	Type[1]	Comparison	Dependent measures[2]	Simulator aim	Sample (*N*, Age)	Real drive	Simulator characteristics and fidelity[3]
Abdel-Aty et al. (2006)	R	Between	S	Validate the driving simulator for speed and safety.	N_{sim} = 58–62 per scenario N_{real} = 420 observations	Radar record of regular road users	UCF driving simulator; I-Sim Mark-III system. Full vehicular cabin with automatic transmission on motion platform with 6 degrees of motion. Five channels (front, side, and back) displaying integrated audio and 180° FOV visual environment and console. **Fidelity score:** 14 *HIGH* V = 4 M = 5 P = 5
Alm (1996)	R	Within	S, L, O	Confirm if lateral position differs between sim and real drive.	*N* = 17 (9F) 27–46 years	Instrumented vehicle of same make and model as sim (SAAB 9000)	VTI driving simulator; 120° FOV in a SAAB 9000 cabin. Simulator sits on motion base. **Fidelity score:** 11 *MEDIUM* V = 3 M = 3 P = 5
Barker et al. (1978)	R	Within	D/E	Identify significant difference between results of Oklahoma driving road test and simulator test	*N* = 900	Participants completing a practical driving test, no vehicle description.	Oklahoma Driver Screening Simulator; vehicle cab, single screen; no detailed description provided. **Fidelity score:** 7 *LOW* V = 1 M = 1 P = 5
Bédard et al. (2010)	J	Within	D/E	Validity of driving simulator	*N* = 8 67–81 years	Information not provided	STISIM programming, with 135° FOV across three computer monitors. No information provided about the physical format. **Fidelity score:** 7 *LOW* V = 3 M = 1 P = 3
Bella (2005)	J	Between	S	Validation of speed variable for CRISS simulator	N_{real} = 636 vehicles N_{sim} = 35	Naturalistic observation: laser speed measures or video camera of 8 sites.	CRISS driving simulator; interactive simulator using Vehicle Dynamic Analysis Nonlinear, with a static base. Facilitated by a complete vehicular cabin with integrated pedals and console. Visual projection on three projectors giving a 135° FOV. **Fidelity score:** 9 *MEDIUM* V = 3 M = 1 P = 5
Bella (2008)	J	Between	S	Validation of simulator for two-lane rural road	N_{real} ≈ 2000 recordings N_{sim} = 40 23–60 years	Laser speed measure	CRISS driving simulator; as in Bella (2005) **Fidelity score:** 9 *MEDIUM*
Bham et al. (2014)	J	Between	S	Simulator validation	N_{sim} = 46 19–53 years	Speed measures using video detection system	Fixed base, vehicular cabin with integrated functioning controls and pedals; visual environment displayed via three projectors onto a single curved screen giving 120° FOV. **Fidelity score:** 9 *MEDIUM* V = 3 M = 1 P = 5

Reference	Type[1]	Comparison	Dependent measures[2]	Simulator aim	Sample (N, Age)	Real drive	Simulator characteristics and fidelity[3]
Blaauw (1982)	J	Within	S, L	Simulator validation	$N = 48$ 18–36 years	Instrumented vehicle – ICARUS (Blaauw and Burrij, 1980)	Fixed base simulator with three small screens giving 120° FOV. Vehicle cabin is a mock-up of the instrumented vehicle. **Fidelity score:** 7 *LOW* V = 3 M = 1 P = 3
Branzi et al. (2017)	J	Between	S	Simulator validation for speed variable at urban crossings	$N_{real} = 39–189$ observations per zone $N_{sim} = 34$ 24–65 years	Speed recorded via laser speed meter at 21 sites along a 2.5km stretch of road.	LaSIS driving simulator; complete integrated vehicle cabin on motion platform, four projectors display visual environment on 200° cylindrical screen, LCD monitors for rear mirrors **Fidelity score:** 15 *HIGH* V = 5 M = 5 P = 5
Carter & Laya (1998)	B	Within	S, O	Investigate the effect of experience, task, and situation on visual search strategies	$N = 16$ aged 30–50 years $N = 8$ aged 18–21 years	Instrumented vehicle on closed track, no additional information provided	Fixed base simulator with ‘arcade controls’ using Silicon graphics onto a single projector screen. **Fidelity score:** 6 *LOW* V = 2 M = 1 P = 3
Daurat et al. (2013)	J	Within	L, O	Degree to which the impairment of benzodiazepines can be measured in simulated driving environment.	$N = 16$ 25–35 years	Dual control vehicle fitted with lateral position recorder. No additional information provided.	Fixed based, fully equipped Citroen C2 adapted by Oktal. Simulated environment from LEPSIS projected onto a single forward and rear screen giving 120° forward FOV and rear view via mirrors. **Fidelity score:** 9 *MEDIUM* V = 3 M = 1 P = 5
Davenne et al. (2012)	J	Between	L	Reliability of simulator in fatigue research; impact of sleepiness and fatigue at the wheel in simulator and real driving conditions	$N_{real} = 14$ $N_{sim} = 20$ Age $M = 22.3$ years	Vehicle driven on highway as part of a separate study (Sagaspe et al. 2008, as citied in Davenne et al. 2012). No description provided.	Fixed based INRETS-MSIS SIM2 simulator using a computer screen and video game steering wheel. **Fidelity score:** 4 *LOW* V = 1 M = 1 P = 2

Reference	Type[1]	Comparison	Dependent measures[2]	Simulator aim	Sample (N, Age)	Real drive	Simulator characteristics and fidelity[3]
Edwards et al. (1977)	J	Within	D/E	Compare OR driving performance with two simulators	N_{real} = 304 N_{sim} = 200	Observed by raters in their taxi	Allstate Good Driver Trainer; vehicle cabin with functioning integrated controls and pedals; fixed base. Visual environment is projected onto single screen. Also incorporates a complex error scoring system. **Fidelity score:** 8 *MEDIUM* V = 2 M = 1 P = 5 Atena Drivotron; vehicle cabin with functioning integrated controls and pedals. Visual environment is projected onto single screen. Cabin swings as steering wheel is turned. Changes in acceleration alters speed and volume of video and engine **Fidelity score:** 10 *MEDIUM* V = 2 M = 3 P = 5
Fildes et al. (1997)	R	Between	S	Validation of simulator	N_{real} = 24 22–52 years N_{sim} = 24 22–40 years	Instrumented vehicle	MUARC driving simulator; four projector screens giving 180° FOV in front and 1 rear screen, motion based with road sensation feedback and full vehicular cabin. **Fidelity score:** 15 *HIGH* V = 5 M = 5 P = 5
Fors et al. (2013)	R	Within	S, L, O	Simulator validation, particularly interested in detection of driver fatigue	N = 16 20–55 years	Instrumented Volvo XC70 with dual controls	VTI driving simulator III; moving base simulator with a 120° forward FOV. Full integrated cabin of Saab 9-3 **Fidelity score:** 13 *HIGH* V = 3 M = 5 P = 5
Freund et al. (2002)	J	Within	D/E	Pilot validation of simulator for older cognitively impaired adults	N = 9 (4 cognitively impaired) 67–78 years	Dual brake equipped vehicle	STISIM Drive; no description of the simulator provided. **Fidelity score:** 7 *LOW* V = 3 M = 1 P = 3
Galski et al. (1992)	J	Within	D/E	Driving simulator performance as a relate to OR performance	N = 35 18–87 years M = 1.8yrs since traumatic head injury	OR and ‘lot’ driving course; no vehicle information provided	Doron L225 Driving System; no description of the simulator provided. **Fidelity score:** 5 *LOW* V = 1 M = 1 P = 3
Hallvig et al. (2013)	J	Within	S, L	Compare real and simulated driving under extended wakefulness	N = 10 Age M = 40 years	Dual control Volvo S80; drive was 55km loop of which a subsection appears in the simulator scenario.	Advanced motion-based simulator, with partial cabin of a Volvo 850. Three channels of visual environment gave 120° FOV. **Fidelity score:** 12 *HIGH* V = 3 M = 5 P = 4
Harms (1996)	B	Within	S, L	Behavioural validity of the VTI Driving Simulator	N = 7 24–54 years	Instrumented vehicle similar to simulator	VTI driving simulator; as outlined in Al (1996) **Fidelity score:** 11 *MEDIUM*

Reference	Type[1]	Comparison	Dependent measures[2]	Simulator aim	Sample (N, Age)	Real drive	Simulator characteristics and fidelity[3]
Helland et al. (2013)	J	Mixed	L	Establish a simulator test battery for the evaluation of ethanol and validate this battery when compared with real driving	N_{real} = 10 N_{sim} =18 25–35 years	Auto instrumented car (Volvo V70 2.4s) with dual controls	Full vehicular cabin of common car with integrated controls and motion base. Large screens gave 180° forward and 90° rear FOV with internal and external mirrors. **Fidelity score:** 15 *HIGH* V = 5 M = 5 P = 5
Hou et al. (2014)	J	Within	O	Integration of traffic and driving simulator protocol; validation	N = 15 21–39 years	Participants in own vehicle, vehicle equipped with on-board diagnostic device	Front seat vehicular cabin positioned on full motion platform. Four screens project visual environment giving forward and rear FOV. **Fidelity score:** 14 *HIGH* V = 4 M = 5 P = 5
Hulme & Thorpe (2013)	R	Within	D/E	Validation of simulator of elderly driver evaluations	N = 10 65–77 years	Dual equipped and instrumented 2008 Ford Taurus.	<u>UB driving simulator</u>; no detailed description provided. Appears to have large FOV on projector and partial vehicular cabin. **Fidelity score:** 10 *MEDIUM* V = 5 M = 1 P = 4
Johnson et al. (2011)	J	Within	O	Compare OR and simulator physiological responses	N = 9 20–47 years	Drove own vehicle on test route for obtaining licence	Partial vehicular, integrated console and pedals; visual environment by STISIM across three 17-inch monitors giving 135° FOV, side and rear mirrors included on screens **Fidelity score:** 8 *MEDIUM* V = 3 M = 1 P = 4
Klee et al. (1999)	J	Within	S	Simulator validation	N = 30 17–60 years	Instrumented vehicle	<u>UCF driving simulator</u>; Complete vehicle with engine removed and wrap around projector screen from three projectors creating 160° FOV **Fidelity score:** 10 *MEDIUM* V = 4 M = 1 P = 5
Lee et al. (2003)	J	Within	D/E	Validate a driving simulator for an OR test for older adult drivers.	N = 129 60–88 years	Own vehicle	STISIM programmed simulator with single small PC monitor and with steering wheel control **Fidelity score:** 5 *LOW* V = 1 M = 1 P = 3
Lee et al. (2002)	R	Between	O	Compare braking responses in simulator and test track	N_{sim} = 16 25–55 years	N/A. On-road drive used archival data and information was not provided.	<u>Iowa driving simulator</u>; fully instrumented vehicle cabin on a full motion-based platform; multiple projectors give drivers a 190° forward and 60° rear FOV **Fidelity score:** 14 *HIGH* V = 4 M = 5 P = 5

Reference	Type[1]	Comparison	Dependent measures[2]	Simulator aim	Sample (N, Age)	Real drive	Simulator characteristics and fidelity[3]
Lee et al. (2013)	R	Between	S	Comparisons of driver speed across four simulators when compared with data collected on-road	N_{NADS} = 48 N_{FHWA} = 48 N_{WTI} = 48 $N_{miniSim}$ = 23 N_{real} = 16 25–45 years	Observational speed data of roads modelled in simulator	NADS; full vehicle cabin within a dome that fully encapsulates the sim. Projecting 360° FOV on the walls of the dome, that is mounted on full a motion platform. **Fidelity score:** 15 *HIGH* V = 5 M = 5 P = 5 FHWA; full vehicle cabin mounted on a motion platform with 3 degrees of motion; 240° FOV projected onto large screens. **Fidelity score:** 12 *HIGH* V = 4 M = 3 P = 5 WTI; full vehicle cabin mounted on a motion platform with 3 degrees of motion; 240° FOV projected onto large screens; and side mirrors replaced by smaller screens. **Fidelity score:** 14 *HIGH* V = 4 M = 5 P = 5 NADS miniSim; no motion base; arcade style simulation with seat and steering wheel from an actual vehicle. Three TV screens give approximately 135° FOV **Fidelity score:** 7 *LOW* V = 3 M = 1 P = 3
Li et al. (2013)	J	Within	O	Validating simulator for physiological signals	N = 15 Age M = 25 years	VW polo; similar car to simulator	Autosim driving simulator; full static integrated vehicle cabin, three projector screens give 120° FOV. **Fidelity score:** 9 *MEDIUM* V = 3 M = 1 P = 5
Mayhew et al. (2011)	J	Within	D/E	Validation of simulator to OR driving and between various driving groups	N = 64 (57 complete) Age M = 16 years	Cars supplied by an insurance company from a licencing office	Interactive STI generated simulation; comprised of three screens and single 'arcade style' physical setting; no motion base. **Fidelity score:** 7 *LOW* V = 3 M = 1 P = 3
McAvoy et al. (2007)	J	Between	S	Validate simulator and the effectiveness of traffic control devices during night-time driving	*Real Drive (per site)* N_{test} = 320–451 $N_{control}$ = 200 N_{sim} = 127	Field observations of work zones on 6 freeways (4 test and 2 control)	Five computer screens present 225° panoramic FOV; no motion base. **Fidelity score:** 8 *MEDIUM* V = 4 M = 1 P = 3
Meuleners & Fraser (2015)	J	Within	L, D/E	Compare driving performance across simulator and OR drive for errors	N = 47 18–69 years	Instrumented vehicle with dual controls	UC-win/Road Simulator; full seat of automatic car with seat belt, pedals, steering wheel, ignition and key, and speedometer. On fixed base, in front of three monitors giving 180° FOV. **Fidelity score:** 8 *MEDIUM* V = 4 M = 1 P = 3

Reference	Type[1]	Comparison	Dependent measures[2]	Simulator aim	Sample (N, Age)	Real drive	Simulator characteristics and fidelity[3]
Mueller (2015)	T	Within	O	Validate response to increased workload in simulator to increased workload in real drive	$N = 34$ 25–37 years	2009 Chevy Impala instrumented vehicle	WTI; outlined in Lee et al. (2013) **Fidelity score:** 14 *HIGH*
Philip et al. (2005)	J	Within	L	Are the effects of fatigue, sleepiness, and performance the same in a simulator as on the road	$N = 12$ 19–24 years	Dual control vehicle fitted with camera to record the road	Divided attention steering simulator; Software reproduces a winding road, Ps must keep front of car in middle of lane. Single PC screen; steering wheel. **Fidelity score:** 4 *LOW* V = 1 M = 1 P = 2
Reed & Green (1999)	J	Within	S, L, O	Validation of driving simulator	$N = 12$ 20–30 and 60+ years	Instrumented vehicle: 1991 Honda Accord LX	Integrated vehicular cabin; visualisation constructed with UMTRI software and is projected onto a single large screen with a 33° FOV **Fidelity score:** 8 *MEDIUM* V = 2 M = 1 P = 5
Risto & Martens (2014)	J	Within	O	Does headway choice differ between OR and simulator	$N = 22$ 27–64 years	Instrumented vehicle; Toyota Prius	Skeletal car cabin constructed in front of visual screen giving participants a 180° FOV; fixed base with integrated pedals and speedometer. **Fidelity score:** 10 *MEDIUM* V = 5 M = 1 P = 4
Santos et al. (2005)	J	Between	S	Comparison between low-medium virtual systems, simulator, and instrumented vehicle for driving impairment by secondary task	$N_{lab} = 24$ Age $M = 29$ years $N_{sim} = 24$ Age $M = 32$ years $N_{real} = 24$ Age $M = 40$ years	Instrumented vehicle, Renault 19 with dual controls.	Laboratory Driving simulator; no motion base, single PC screen, 'arcade style' **Fidelity score:** 5 *LOW* V = 1 M = 1 P = 5 Leeds Driving simulator; no motion base, complete vehicular cabin. Visualisation projected onto 2.5m cylindrical screen with 230° FOV and rear projection screen of 60° FOV seen in rear vision mirror **Fidelity score:** 11 *MEDIUM* V = 5 M = 1 P = 5
Senserrick et al. (2007)	J	Within	S	Validation of free speed in simulator and OR	$N = 21$ 16.0–16.6 years	Data logger in their own vehicle	NADS; as outlined in Lee et al. (2017) **Fidelity score:** 15 *HIGH*
Shechtman et al. (2009)	J	Within	D/E	Determine if error type and error total are similar in simulator and OR	$N = 39$ 25–45 and 65–85 years	Dual control instrumented vehicle.	Simulator scenario developed with STISTIM integrated with full vehicle, visual environment projected onto large projectors for 180° FOV. No motion base. **Fidelity score:** 10 *MEDIUM* V = 4 M = 1 P = 5

Reference	Type[1]	Comparison	Dependent measures[2]	Simulator aim	Sample (*N*, Age)	Real drive	Simulator characteristics and fidelity[3]
Törnros (1998)	J	Within	S, L	Validation of driving behaviour in simulated and OR tunnel	$N = 20$ 23–52 years	Instrumented SAAB 9000 CDE	VTI driving simulator; as outlined in Alm (1996) **Fidelity score:** 11 *MEDIUM*
Veldstra et al. (2015)	J	Within	L	To determine if the simulator and OR driving generated equal effects of psychoactive substances	$N = 24$ Age $M = 23.6$ years	Instrumented vehicle, other details not provided	Fixed-base mock-up of a car with standard controls linked to computer with ST software simulator. Visual environment displayed 210° FOV **Fidelity score:** 8 *MEDIUM* V = 4 M = 1 P = 3
Wade & Hammond (1998)	R	Within	L	Determine reliability, and validity of driving performance in wrap-around simulator	$N = 14$ 18–34 years	Instrumented Honda Accord	HFRL simulator; detailed description not provided, referred to as "mid-level" by authors. Provided details: "wrap-around" visual FOV and static **Fidelity score:** 7 *LOW* V = 3 M = 1 P = 3
Wang et al. (2010)	J	Between	S, L	Investigate the differences between in-vehicle information interfaces and validate the simulator for this	$N_{real} = 28$ Age $M = 23.9$ years $N_{sim} = 30$ Age $M = 24.3$ years	Instrumented vehicle	MIT driving simulator; fixed base, full vehicular cabin with full integration of console and pedals. Visual environment (STISIM) projected onto a single screen with 40° FOV with rear vision mirror in screen. **Fidelity score:** 8 *MEDIUM* V = 2 M = 1 P = 5
Yun et al. (2017)	J	Between	L	Validation of simulator for lane change behaviours in urban environment	N_{real} = data not provided $N_{sim} = 80$	Information not provided	Fully instrumented vehicle cabin on motion base platform. Immersive projector with 250° FOV forward, monitors for side and rear mirrors **Fidelity score:** 15 *HIGH* V = 5 M = 5 P = 5
Zöller et al. (in press)	J	Within	O	Validity of braking behaviour in urban intersections	$N = 42$ Age $M = 29.3$ years	Instrumented automatic vehicle	Arcade style physical layout with full motion platform (hexapod). Visual environment presented on three computer monitors with 60° or 180° FOV. FOV and motion (on/off) were manipulated → 4 fidelity scores: **1. Fidelity score:** 7 *LOW* V = 3 M = 1 P = 3 **2. Fidelity score:** 8 *MEDIUM* V = 4 M = 1 P = 3 **3. Fidelity score:** 9 *MEDIUM* V = 3 M = 3 P = 3 **4. Fidelity score:** 10 *MEDIUM* V = 4 M = 3 P = 3

Notes.

[1] Type classification: **B**ook section; **J**ournal article; **R**eport; **T**hesis
[2] Dependent measures: **S** speed or speed variation; **L** lane/lateral position or lane/lateral position variation, or lane crossing; **D/E** overall driving score or errors; **O** other (e.g., eye tracking, EEG, reaction time, headway, steering).
[3] Fidelity score; sum of **V**isual, **M**otion, and **P**hysical fidelity (see Table 4)