5

Twelve Practical and Useful Questions About Driving Simulation

Jeff K. Caird
University of Calgary

William J. Horrey
Liberty Mutual Research Institute for Safety

Abstract

The Problem. A number of practical questions or concerns are frequently expressed about driving simulators. For example, what can you use a driving simulator for? What are the advantages and disadvantages of using a simulator? Probably the most important question though is whether driver behavior in a simulator mimics that which is exhibited while driving in the real world. These and other epistemological questions are discussed. ***Role of Driving Simulators.*** The first driving simulator research paper appeared in 1934. Since then, driving simulators have evolved into a flexible means to measure a variety of variables while drivers drive in a wide range of traffic environments. Empirically, driving simulation allows researchers to exert control over confounding variables that are common in actual driving. Testing of drivers in crash-likely conditions is also possible, which for ethical reasons cannot be done in the real world. Uses of driving simulators for research are discussed. ***Key Results of Driving Simulator Studies.*** Some have claimed that driving simulator research has not contributed to the progression of knowledge about driving performance, behavior, or safety. An inspection of the chapters in this *handbook* attests to the scope and scale of contributions. Further, the corpus of research using simulation has been constant and is now increasing. Areas of contributions are analyzed. ***Scenarios and Dependent Variables.*** The selection of scenarios that address certain research questions is dependent on the joint capability of a simulator and research team, which often evolves with experience. A list of common dependent variables is provided. ***Platform Specificity and Equipment Limitations.*** A realistic appraisal of a simulator's capabilities in light of the results produced is required, including limitations therein.

5.1 Introduction

A number of commonly and not so commonly asked questions about driving simulators are here addressed. Some of the questions are relatively easy to answer, whereas others do not necessarily have an answer. Other questions are long-standing criticisms of driving simulation (e.g., Dingus et al., 2006; Evans, 2004) while others have yet been overlooked. The answers, in some cases, are a list of pointers to others who have addressed the criticism or have a solution. Specifically, the following questions are addressed:

1. What are the origins of driving simulation?
2. What can a driving simulator be used for?
3. What advantages and disadvantages are there in using a driving simulator?
4. Where are driving simulators that are used for research located throughout the world?
5. When using driving simulators, what special considerations should be discussed when applying to an ethics committee or institutional review board (IRB)?
6. What dependent variables can be collected using a driving simulator?

7. How do you map the necessary realism or fidelity of a driving simulator to the specific research questions?
8. To what degree are the results from a simulator similar to those measured in the real world?
9. How well do the results from simulator, on-road and more basic laboratory studies correspond with one another?
10. What are common threats to internal and external validity?
11. Is there evidence that driving simulators have contributed to the advancement of knowledge since the 1970s?
12. Based on the progress of software, hardware and projection capabilities, where will the future of driving simulation lead?

The first questions deal with historical information, the practical uses of driving simulators and the prevalence of simulators today. The remaining questions, which parallel the progression of research activities in a study, focus on critical issues, concerns and threats in using simulators in scientific endeavors. While simulators carry tremendous flexibility and potential utility, we believe that it is important that all users of simulators be cognizant of the inherent issues and limitations.

5.2 The Twelve Questions

1) What are the origins of driving simulation?

Driving simulators have been used to explore aspects of driving since the 1960s (see this book, chap. 2 by Allen, Rosenthal, & Cook, "A Short History of Driving Simulation"). While this epoch represents the period in which they became more prevalent, the origin of the concept of driving simulation dates much further back. For many years, the static simulator, shown in Figure 5.1, developed by De Silva (1936) was considered one of the potential originators of driving simulation. For instance, Gibson and Crooks (1938) cite De Silva's 1935 research at the State College of Massachusetts in their seminal paper on a theory of driving.

The first author (JKC), while engaged in a game of scholarship, tracked an even earlier reference to driving simulation. This was the result of a challenge issued by Hancock & Sheridan (authors of "The Future of Driving Simulation" this book, chap. 4) some 20 years ago to find the original or first citation pertaining to driving simulation. (It is worth mentioning to the current generation of would-be scholars that library stacks are not searchable like e-journals, at least not until all known publications are scanned and accessible.) The reference for perhaps the first driving simulator is shown in Figure 5.2. This 75 year-old citation has not been shown for a number of generations until now.

Miles and Vincent built an apparatus to create the illusion of driving circa 1934. The purpose of the driving apparatus was to develop compulsory tests for drivers to combat the public indignation about the high rates of road traffic deaths and injuries. The specifics of the device are as follows:

> An apparatus has been constructed at the Institute [U.K.] which gives a very strong illusion of actual driving, in which the conditions can be kept absolutely constant, and which keeps a printed record of the track and speed of the driver. The subject sits in a dummy car, with the usual controls and 'drives' (see [the figure; here Figure 5.2]). The illusion of movement is given by a moving picture on the projector mounted on the chassis of the dummy car. The motor driving the projector is mounted on the chassis of the dummy car, and the vibration it causes varies with the apparent speed, adding greatly to the illusion. The driver can go where he [or she]

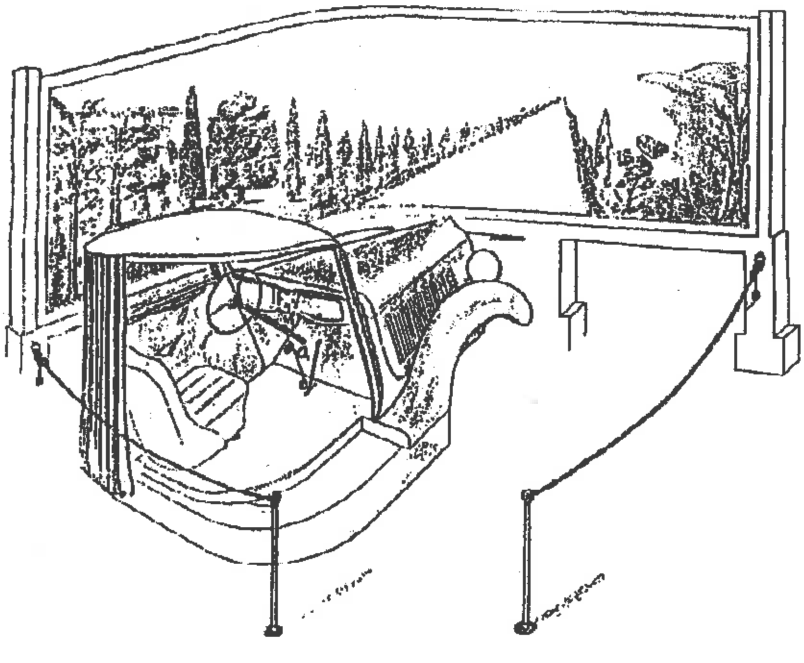

FIGURE 5.1 An early device that resembled a driving simulator appeared in the journal *The Human Factor* in 1936 (De Silva). The mural was a static scene.

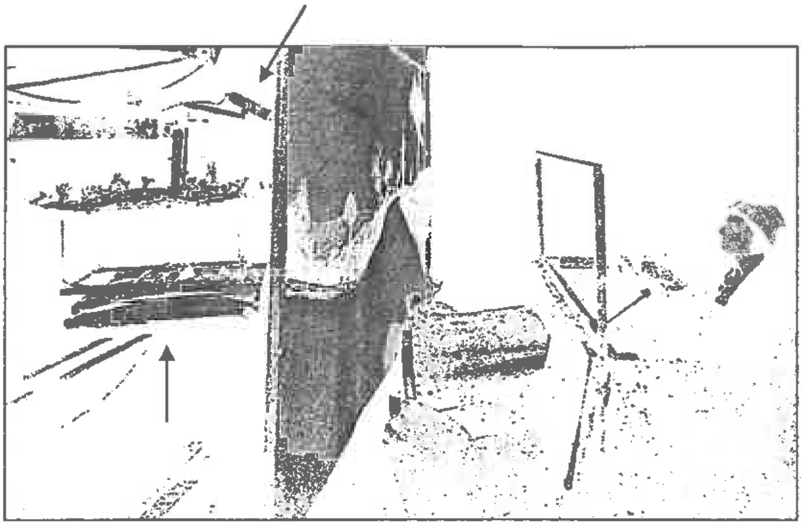

FIGURE 5.2 The earliest citation of a system that resembled simulated driving appeared in the journal *The Human Factor* in 1934 (Miles & Vincent). This journal was published in the U.K. for about 10 years during the 1920s and 1930s. The bottom arrow points to a set of miniature models that were projected by a device (top arrow) onto the screen before a driver.

likes—on the wrong side of the road, on the grass, into the hedge—he can take any turning he likes or he can swing his car around, if the road is wide enough, and go back the way he came. Behind the screen is a miniature landscape with roads an inch or so wide and trees and hedges on the same scale. A projector passes over the roads of this model, throwing an enlarged moving picture upon the screen. The observer sees just the view he would get if he were actually driving his car along the roadway. At the same time a pen is tracing out the exact track on a plan. The track of the pen is a dotted line, the dots being made at the rate of ten per second. It is thus possible to calculate the speed of any part of the track by measuring the distance between the dots. The subject is first allowed ten minutes to accustom himself to the 'feel' of the apparatus, and then a number of track records are taken (Miles & Vincent, 1934, pp. 253–255, brackets [inserted information]).

The presentation of miniature models continued to be used for many years (cf., see Allen, Rosenthal, & Cook, this book, chap. 2; Blaauw, 1982; Sheridan, 1970; Wierwille & Fung, 1975). In addition to predating De Silva's device (Figure 5.1), Miles and Vincent built a driving apparatus that was interactive. With De Silva's device in Figure 5.1, the imagination of the driver would have been required to produce a sense of driving while viewing the static mural. Today we still tacitly ask our participants to make believe that they are driving even though they are not. As an operational assumption, researchers generally accept that drivers believe they are driving when using a simulator, although this is open to considerable debate.

To place the "driving apparatuses" of Miles and Vincent and De Silva into historical context, the Link flight simulator was introduced in 1929 and put into extensive use during WWII as a means to train pilots to fly by instrumentation (in Wikipedia, search on *<Flight Simulation>*, and Web Figure 5.1). Interestingly, simulator sickness was prevalent when using the Link trainer, which is illustrated in a cartoon from 1942 (Web Figure 5.2). Flying and driving are very different constellations of tasks. Pilots are a highly specialized, homogenous group (e.g., Lee, 2005; Tufano, 1997). There is no analog in driving for instrument flight rules (IFR), where an aircraft is piloted using the instrument panel alone (i.e., no visual information from outside the cockpit). A large, heterogeneous population of drivers use complex visual information to control their vehicles, which was first theoretically described by Gibson and Crooks (1938) and extended by Schiff and Arnone (1995). As such, a mass-produced, easily accessible vehicle made the need to train in a surrogate environment largely irrelevant for economic, task and access reasons.

2) What can a driving simulator be used for?

Many researchers and practitioners have a general idea about what a driving simulator can be used for. We compiled a comprehensive list of actual and possible uses of driving simulators:

- Perform research on traffic safety.
- Conduct a structured training curriculum.
- Assess training and education effectiveness.
- Assess those who are about to be licensed or require re-licensing.
- Evaluate vehicle designs.
- Test traffic control devices.
- Understand the effects of basic human limitations on driving.
- Examine the sources and consequences of driver impairments.
- Develop and evaluate new in-vehicle and co-operative infrastructure technologies.
- Assess roadway geometries before they are built.
- Play a game.
- Develop the software and hardware skills of students and technicians.

- Determine the capability of the simulator technology.
- Visualize crashes and crash contributors.
- Understand the effects of disease conditions and various medications on driving.
- Demonstrate the advanced technology of a country, institution or corporation.

The details of each of these uses are more complicated than what a simple list can convey and go beyond the desired scope of this chapter. Fortunately, many of these topics are eloquently grouped and described in other chapters in this *Handbook* (see Table of Contents). However, one topic in particular is often overlooked; namely, the last one on the list. Driving simulators are used for public relations purposes. For instance, doing demonstrations, or demos, is an activity that both of this chapter's authors, along with their colleagues and their students have done for many years for the University of Calgary (and the University of Minnesota before that) and the University of Illinois. I (JKC) have felt an obligation to talk about the research that we do and why it is important to society because we receive public funding. In fact, the mission of those associated with the University of Calgary Driving Simulator (UCDS) is to reduce traffic fatalities and injuries through the creation and dissemination of basic and applied knowledge. Dissemination of knowledge through peer-reviewed publications is only one of many ways to teach researchers, allied professions, the community, students, and drivers. In addition, many researchers who get into simulation are unprepared for the demands placed on them to perform activities that are seemingly unrelated to research.

Driving simulators naturally attract interest from people of all ages. Collectively, we have given well over 100 formal demos over the past eight years to: government officials, media, corporate heads, engineers of all stripes, student organizations, high school classes, class laboratories (e.g., Introduction to Human Factors), professional organizations, conferences, donors, and other researchers and scientists (including Nobel Laureates). Group size matters. For everyone to be able to try the simulator and ask dozens of questions in about an hour, a group of 8 to 12 people is ideal. Restricted laboratory space made one visit from a group of 30 important persons into a circus.

Once a group comes, we try to teach our visitors something about research, traffic safety and driving simulation before we let them go for a drive. The message is often tuned to the level of the audience. After a 5 to 10 minute slide show about the research activity done with the UCDS, visitors take turns, if they would like to, traveling through a demonstration drive that we developed for this purpose. Most of the demonstration time is unstructured while each visitor gets an opportunity to try out the simulator. While this is going on, we try to answer all the questions that we get, which come quickly once our visitors see and experience the possibilities. Our enthusiasm to do this public service, which is desired by institutions but not rewarded or recognized, has waned somewhat over the years. Saying no to requests, especially while studies are underway, is more common. Contrary to conventional wisdom, demonstrations have a low relationship to funding opportunities.

3) What advantages and disadvantages are there in using a driving simulator?

A number of known advantages and disadvantages about driving simulators are shown in Table 5.1. Arguments marked with an asterisk * are expanded or debated in greater depth in Questions 5 and 11, respectively. Disadvantages marked with a ‡

TABLE 5.1 Advantages and Disadvantages of Driving Simulators

Advantage	Disadvantage
• *Has the capability to place drivers into crash likely situations without harming them, such as when they are using drugs, fatigued, engaging in police pursuits, during extreme weather, using new technologies, among other dangerous activities.	• *Simulated crashes do not have the same consequences as a real crash and may affect subsequent behavior. Crashes in a simulator may have an unknown psychological impact on participants.
• Many confounding variables that occur in on-road driving can be controlled when driving simulation is used (e.g., weather, traffic, lighting, frequency of vulnerable road users, wind, potholes, proportion of vehicle types, irrational or unexpected behavior of other drivers, and so forth).	• These confounding or interacting variables that occur in the real world also need to be understood and, since they cannot be fully recreated in simulators, are not necessarily amenable to testing (as yet). In other words, understanding driver behavior is in the interacting details.
• All of the sensory details of the real world are not used by drivers anyway. Perceptual information (Gibson, 1986) for driving is knowable and can be faithfully reproduced using simulators.	• The real world can never be perfectly reproduced (for now). The important combinations of real-world information and feedback that are important to driving are not completely known.
• Events or scenarios can be identically repeated for each participant.	• ‡Each exposure or trial affects responses to subsequent exposures.
• Simulators offer cost savings through flexible configurability so that a wide range of research questions can be addressed (see Jamson, this book, chap. 12).	• High-end simulators, such as NADs, require considerable hardware and software development to address a limited number of research questions.
• Even low-cost, low-fidelity simulators in the right hands can address a wide variety of interesting research questions.	• Low-cost simulators can be imprecise and inflexible and therefore do not address all needs.
• Driving simulation is compelling and elicits emotional reactions from drivers that are similar to those of actual driving.	• Drivers do not believe in the authenticity of the simulation at a fundamental level and responses are based on this perception.
• *Simulators are good at assessing *driver performance* or what a driver *can do* (Evans, 2004).	• *Simulators are not able to address questions of *driver behavior*, which is what a driver does do in their own vehicle (Evans, 2004).
• A structured driver training curricula can be set up and run for new drivers and for some skills, transfers to the open road (see Pollatsek et al., this book, chap. 30)	• The extent that the driver training transfers to on-road skills is not known nor is the relative cost-effectiveness of such programs (see Jamson, this book, chap. 12).

are addressed in Question 10, which deals with common threats to internal and external validity.

To justify using a simulator, researchers frequently mention the advantages of doing so over other methods such as naturalistic observation, instrumented vehicles or laboratory studies. (And the same is true of using other methods over driving simulation.) Often the counterargument is overlooked, but in Table 5.1 the advantages and contrasting views are paired together. The use of "advantage" and "disadvantage" only loosely applies to the conceptual comparisons. The intent of providing contrasting views is to encourage critical thought about driving simulators and alternate methods. Even the most capable and proficient researchers cannot offset all of the potential disadvantages in conducting simulator research; however, we believe that all researchers and users of simulators should *at least* be cognizant of these issues.

One obvious way to use Table 5.1 is to emphasize the disadvantages of driving simulation to justify pursing an alternative research agenda such as naturalistic observation. Emphasizing the disadvantages of a methodological approach and ignoring the advantages (or vice versa) is a form of self-deception that may be practiced because of the immense efforts required to build, operate and effectively use expensive and complex methods. However, the abstract relationships among the advantages and disadvantages and the subtleties of each statement require a much deeper consideration that transcends simplistic scorecard thinking.

4) Where are driving simulators that are used for research located throughout the world?

A partial list of research simulators throughout the world is kept by INRETS (French National Institute for Transport and Safety Research, http://www.inrets.fr/ur/sara/Pg_simus_e.html). The list of countries with research simulators includes Australia, Canada, France, Germany, Japan, China, The Netherlands, New Zealand, Poland, Spain, Sweden, the U.K., and the U.S.

All things considered, checking out who has what can lead to simulator envy. That being said, the hardware and software used to create driving simulation is less than one-half of the capital necessary to create research "excellence and innovation". Researchers, students and technicians comprise the most important ingredients of a successful simulation facility. Inspection of representative publications from various facilities should moderate initial impressions of various kinds of interesting driving simulators. The authors in this *Handbook* represent an obvious starting place to establish who is known for what.

5) When using driving simulators, what special considerations should be discussed when applying to an ethics committee or institutional review board (IRB)?

To perform studies with human participants requires an ethics certification. Each study, and substantive modification, requires interaction with an IRB or Ethics Committee. A typical list of issues that need to be addressed when an ethics application is submitted includes: Informed consent, consent to use participant image, negative transfer, simulator sickness, parental consent (if needed), screening criteria, recruiting procedures, questionnaires, experimental script or verbal protocol, payment, and debriefing. Many of these considerations, including simulation sickness, are discussed by Trick and Caird, this book, chap. 26. However, negative transfer and use of participant images are not.

Negative transfer may occur after a driver has been exposed to driving in a consequence-free environment (also one that may not faithfully replicate real-world driving conditions). Following the experiment, he or she may not be adequately (or fully) recalibrated to real-world driving and may operate his or her vehicle with the same disregard for consequences. At best, the importance and impact of negative transfer from simulation is an empirical question that remains unresolved. In aviation, pilots are required, in some companies but not others, to wait 24 hours after using a flight simulator before flying a plane. However, the empirical basis for the rule, despite years of research, has not been found.

This topic is not often discussed in the context of simulation research or IRB protocol. In general, it is advisable to err on the side of caution and keep participants in the laboratory for a period of time following their involvement in the simulator, filling out questionnaires or the like, in order to afford them a longer time to readjust to the real world. In some situations, it may be advisable to provide transportation home for participants.

The use of participant images for presentation purposes, once a study is over, is an extension of informed consent procedures. Images can include still pictures and videos of participants which have been captured during the course of experimentation. Uses of images at conferences include illustrating certain patterns of performance or behavior such as crashes, facial reactions, and interaction with nomadic devices. The principles of full disclosure, choice and confidentiality apply. Practically, we have used a second informed consent form that outlines and illustrates how images will be used and treated confidentially. A participant may choose to be in a study and not have his or her image used. Accidental dissemination of images by sharing presentation slides or videos should be avoided because control over the images is no longer possible. Careful consideration of the illustrative value of using a video clip or picture is also recommended.

6) What dependent variables can be collected in a driving simulator?

A number of researchers convinced us of the importance of a handy dependent variable list for driving simulation. Investigators should have a good idea about the scope of variables that can be collected and interpreted. A number of common questions arise about which dependent variable can be measured and how to interpret each relative to past use. The selection of dependent variables occurs based on prior use, simulation capability, researcher expertise, practical and applied generalities, and theoretical considerations. More succinctly, the choice of dependent variables is made based on the questions being asked. Table 5.2 lists common, but not necessarily agreed upon, groupings of dependent variables. Other measures have been used that are not listed here, so this is not an exhaustive list. In addition, new and interesting measures should be developed, so the purpose of this list is to provide initial guidance and easy reference.

TABLE 5.2 Driving Simulation Dependent Variables, Descriptions and References

Variable Classification	Variable	Description	Sample Reference
Longitudinal Control	Speed	Travel speed in km/h or mph.	HASTE (n.d.); Tijerina, Barickman, & Mazzae (2004)
	Speed Variability	Standard deviation of speed.	
	Time or Distance Headway	Time or distance to the rear bumper of the lead vehicle.	
Reaction Time	Perception Response Time (PRT)	PRT is the time to ease the foot off accelerator and initially press brake.	Green (2000); Lamble, D., Kauranen, T., Laakso, M., & Summala, H. (1999, p. 620); Olson & Farber (2003); Summala (2000)
	Brake Response Time (BRT)	BRT is the time from a hazard appearance to the brake onset.	
	Time to contact (TTC)	$TTC = ((v2 + 2Ad)^{0.5} \wedge - v_r)/A$ where d is distance, vr is relative velocity of the vehicle ahead, and A, is the deceleration of the vehicle ahead.	
Crash	Crash	The boundary of a driver's vehicle overlaps with other vehicles or objects, or the current vehicle control dynamics exceed those allowed by the vehicle equations of motion (e.g., rollover).	Caird et al. (2008); Horrey & Wickens (2007); Strayer & Drews (2004)
Lateral Control	Lateral Position	Continuous location of a vehicle with respect to a lane reference.	Green et al. (2004); HASTE (n.d.); Godthelp, H., Milgram, P., & Blaauw, G. L. (1984); van Winsum, W., Brookhuis, K. A., & de Waard, D. (2000)
	SDLP	Standard deviation of lateral position.	
	Lane exceedances (LANEX)	LANEX is the proportion of time outside of a lane or a frequency count of the number of times the lane threshold is exceeded.	
	Time to lane crossing (TLC)	TLC is the time to cross a lane boundary at current steering angle and speed.	
	Reversal rate (RR)	RR is the number of steering wheel direction changes per time or distance value.	
Eye Movements	Glance	All consecutive fixations on a target plus preceding transitions (i.e., saccades).	Green (2007); ISO (2002); SAE (2000); Fisher et al., (this book, chap. 18); Horrey et al. (2006)
	Eyes-off-road-time	The sum or proportion of all of the time associated with all glances not directed to the road.	
	Fixation	Momentary direction of the pupil's gaze (separated by saccades).	
	Percent Dwell Time (PDT)	Percent dwell time to a given area of interest (AOI; e.g., instrument panel).	
Workload, Subjective	NASA-Task Load Index (TLX)	A multi-dimensional subjective workload measure composed of six sub-scales and overall workload.	Hart & Staveland (1988)
	Rating Scale Mental Effort (RSME)	Paper-and-pencil instrument that measures workload on a continuous unidimensional scale.	Young & Stanton (2005); Zijlstra (1993)
	Situation Awareness Global Assessment Technique (SAGAT)	Stopping operator (or driver) activity to ask questions about dynamic information needs.	Endsley (2000); Jones & Kaber (2005)
	Driving Activity Load Index (DALI)	A modified workload assessment tool specific to driving activities.	Pauzié (2008)
Workload, Physiological	Heart Rate (HR)	Number of beats for a time period (usually one minute).	Brookhuis & de Waard (this book, chap. 17); Mulder, L. J. M., de Waard, D., & Brookhuis, K A. (2005, p. 20–1)
	HR Variability	Standard deviation of HR.	
	Respiration	Breaths per minute.	
	Electroencephalography (EEG)	Evoked potential amplitude measured from central nervous system.	
	Skin Conductance	Electrical resistance of the skin.	
Other Measures	Entropy	Prediction error of vehicle signals.	Boer (2000); HASTE (n.d.); Horrey & Simons (2007)
	Safety Margins	Amount of space that drivers maintain around their vehicle. Includes but is not limited to headway.	
	Navigation	Assessment of driver's wayfinding ability or memory for trip directions.	
	Other higher-order or aggregate measures...		

Simulation studies do not necessarily focus on crashes although traffic safety epidemiology does. Crashes in simulation are of sufficient importance to elaborate. A common aphorism is that no one has died in a simulator crash and a fundamental question is to understand what crashes mean as an ordinal measurement variable. Evans considered simulator crashes as a thought experiment some time ago:

> Consider a make-believe simulator consisting of an actual car, but with the remarkable property that after it crashes a reset button instantly cancels all damage to people and equipment. What experiments could be performed on such make-believe equipment which would increase our basic knowledge about driving? (Evans, 1991, p. 127)

Two related questions can be circumscribed that do not dismiss driving simulation as a method outright, which appear to be the intent of the rhetorical arguments. First, because crashes in driving simulation have few consequences, does driving performance or behavior change towards more risk taking in the simulator? Second, does the absence of consequences necessarily negate results obtained using driving simulators? The first question is an empirically testable question that surprisingly, to our knowledge, has not been addressed. One group is put into a crash and another is not. Measures of performance are compared. Short of having results from this important experiment, in some situations, a simulated crash can be quite compelling—even frightening. For instance, older drivers, when involved in a simulated crash, are often quite concerned about the well-being of the other driver or pedestrian, which are virtual entities, and express concern over being reported to the licensing authorities. Novice drivers, similarly, are concerned that their parents might be informed when a crash occurs. The loss of privileges with a family vehicle seems to underlie the concern.

Crash experience coupled with information about the outcomes of a crash may be an effective way to offset the absence of any physical consequences. For example, once a crash occurs, the simulation can be stopped and drivers can be given information that estimates the extent of their injuries (including fatalities in more severe circumstances) along with estimated repair costs, based on the collision angles and impact velocities. Crashes may also become adverse to drivers in simulators by imposing some form of financial penalty for crash involvement, a strategy that might be particularly effective in university-age samples (see Ranney, this book, chap. 9). However, by replacing the utilitarian goals for driving with alternative incentives and/or punishments, participants may modify their performance so that it is dissimilar to normal or everyday driving. We note that a number of social motivations such as avoiding an embarrassing crash or showing off for others may also be in play in monitored environments, which include simulation and naturalistic studies. Predictable elicitation of the behavior and describing the impact of it when it occurs on crash risk represent important research directions. The impact of video gaming, and driving video games in particular, on driving simulation performance and actual driving is another area that requires systematic research (Fisher, Kubitzki, Guter, & Frey, 2007).

Creating circumstances in which it is exceptionally difficult for drivers to avoid a crash in a simulator is rather trivial. Placing hazards into a driver's path with little space and time to respond is easy to achieve, though it is important that these hazards or critical events are not unavoidable, such that attentive and responsive drivers can safely avoid them. Tuning the constraints of a particular event so that a range of abilities is accommodated requires some pilot testing. Creating an array of difficult events to contend with is the basis of determining the limitations of drivers to respond while engaged in distraction tasks (e.g., Chisholm, Caird, & Lockhart, 2008). If crashes are unavoidable, there may be greater repercussions for the driver in subsequent experimental blocks or trials such as modifying glances and speed to anticipate hazards.

Another important point of discussion is the issue of repeated crash experience in a simulator. While simulators are effective at generating many crash-likely situations (and therefore many crashes), real-world crash exposure or experience is quite low. The extent to which driving simulation crash-likely protocols compress long-term exposure into crashes requires firmer footing in injury epidemiology (Green, 2008). At the individual or practical level, does repeated exposure to crashes in a simulator create unrealistic expectations on the part of drivers? Should a simulator session with a participant be terminated if more than one crash has occurred (see Ranney, this book, chap. 9)? Wickens (2001) discusses the psychology of surprise and the implications for research that tries to employ more than one surprising event for a single participant. Unfortunately, the reality is that you can only truly surprise a research participant a single time (and the experimental setting itself may dampen this prospect at the outset!). Subsequent responses after experiencing a surprising event hasten response times, depending on participant age, by about a half-second (Olson & Sivak, 1986).

Evans (2004, p. 188, *italics* in original) has also thought about the effect of expectancy on reaction times:

> Enthusiasm for driving simulators ignores some of the most basic understanding about the nature of traffic crashes. The discussion above on reaction time showed the primacy of expectancy. Even in experiments using actual instrumented vehicles, reaction times are substantially shorter than in normal driving. Any reliance by traffic engineers on reaction times determined on a simulator no matter how realistic; could produce unfortunate results. However, the reason that simulators are unlikely to produce knowledge relevant to traffic safety is more fundamental than this.
>
> Simulators measure *driving performance*, what the driver *can do*. However, safety is determined primarily by *driver behavior* or what a driver *chooses to do*. It is exceedingly unlikely that a driving simulator can provide useful information on a driver's tendency to speed, drive while intoxicated, run red lights, pay attention to non-driving distractions, or not fasten a safety belt. Twenty-year-olds perform nearly all tasks on simulators better than the 50-year-olds, but it is the 50-year-old who has sharply lower crash risks.

Several important questions are raised in this passage. First, do driving simulators provide reliable reaction time results? Second, can any knowledge about driver behavior be obtained by using a driving simulator? One way to determine if reaction times differ across methods is to compare the results from available research to determine if there are systematic differences. For instance, Caird, Chisholm, Edwards, and Creaser (2007) examined reaction times to yellow lights over the past 50 years. Across studies and methods, there is a close correspondence in reaction time results. Observational, test track, experimental, field and simulator are represented in the collection of studies. This issue is discussed further in Questions 8 and 9.

With respect to drivers' expectations, still others have examined patterns of systematic variance such as when a driver expects or does not expect an event (Olson & Farber, 2003). Green (2000) and Summala (2000) debate the implications on response time for different types of events: unexpected and expected events as well as for surprising intrusions. While one cannot truly surprise a participant more than once in a session, the use of different types of scenarios and event configurations can at least reduce the drivers' ability to anticipate these particular events. A continuum of surprise can be conceptualized depending on the frequency with which events are ordinarily experienced on a day-to-day basis. So, while the participant may quickly come to expect certain types of events, at least they can be made temporally and geometrically uncertain.

Regarding the second issue raised by Evans (2004), is there evidence or methods from driving simulation that shows aspects of driver behavior based on choice or self-paced interaction? Does having a prescribed task to perform necessarily negate results if the rate of a self-paced behavior is not necessarily known? For example, previous research on driver multitasking and distraction often will employ tasks that are outside of the driver's control. That is, the investigators usually prescribe the conditions under which the tasks are performed. While this technique may be useful in examining the interference from concurrent activities, it does not capture the adaptive potential of drivers (see Lee & Strayer, 2004; Horrey & Lesch, 2009 for further discussion). Thus, it is very possible that certain methodologies and—by extension—driving simulation may create artificial situations that are not completely reminiscent of real-world situations. Alternative methods that allow a repertoire of behaviors to occur, such as engaging in distracting activities when stopped at stop lights, are often difficult to categorize and analyze (Caird & Dewar, 2007; Stutts et al., 2005). For example, Yan, Adbel-Aty, Radwan, Wang, and Chilakapati (2008) observed drivers who did not come to a complete stop at an intersection in the simulator and related it to the history of crashes at the actual intersection.

Consideration of behaviors that occur in naturalistic driving studies as a precursor to experimentation would surely reduce the occurrence of simulation studies that are not interesting; when this is done, research should have the potential to change our understanding of driving phenomena. Researchers should also carefully consider, in their discussion of study limitations, how their results may or may not generalize to real-world behavior.

7) How do you map the necessary realism or fidelity of a driving simulator to the specific research questions?

Naïve realism, or the misplaced desire for the precise replication of perceived reality, tends to be the default assumption of lay researchers about many forms of simulation such as medical, flight and driving (Gaba, 2004; Lee, 2005; see Jamson, this book, chap. 12, respectively). Fidelity or realism is the degree to which a driving simulation matches aspects of driving on-road (Meister, 1995). By making a simulation look and feel exactly as real tasks and environments do, one forces a set of requirements onto hardware, software and scenarios. By asking what needs require sufficient fidelity to achieve a certain level of verisimilitude, certain requirements can be relaxed. Within cost constraints, the operational assumption of less experienced researchers is to err on the side of higher levels of fidelity. For example, a signals engineer may be concerned about making sure that the traffic lights in a simulation function according to the correct timing cycles and algorithms. However, the traffic lights may be only relevant when a driver passes through several intersections on a drive. In general, the engineering disciplines tend towards requiring an exactness of the simulation even if these physical and virtual aspects of the simulation are not relevant to the tasks of the driver. Knowing which traffic environment features and tasks are relevant to the driver and to the research question being posed is central to developing an effective experimental design.

The typical approach to solving the mapping between the level of fidelity required in each of the various dimensions and the research agenda is to propose a taxonomy or matrix that lists simulator fidelity in columns and research questions or simulator components as rows. The start of one of these lists is shown in Table 5.3 and is left for the reader to complete (i.e., …). A complete table would describe a general classification scheme for simulators that often is a variant of

TABLE 5.3 Simulation Fidelity by Component or Research Question

	Low	Moderate	High
Moving Base	Fixed	Fixed or limited motion cuing	Moving base
Screen Width	20 degrees	150 degrees	360 degrees
Screen Resolution			
Sign Legibility	Poor	Fair	Good
Night-time Visibility	No capability	Poor	Fair

low, medium and high fidelity. For example, relatively recent pictures of low, medium and high fidelity simulators are provided in Jamson, this book, chap. 12 and Shinar (2007). An excellent discussion of physical fidelity, or the matching of physical components in a simulator with vehicle capabilities, is provided by Greenberg & Blommer, this book, chap. 7. One of the limitations with this matrix approach is that simulator categories migrate from right to left as technology progresses (see Allen, Rosenthal, & Cook, this book, chap. 2). For instance, the graphics of the low-fidelity simulators of today are well beyond the capabilities of driving simulators a decade ago. Knowing where the state of technology is (or was) is important to interpreting the inherent limitations of various studies across time.

The mapping of simulation fidelity to a research question, in part, assumes that a simulator can be prescribed at the time that a particular question is addressed. In reality, the justification for a range of research problems is offered at the time that simulator funding is sought and when the simulator is purchased (see Question 2). Researchers hope that they have anticipated or planned adequately for a research agenda that corresponds to the first few years of activity; that is, if operational costs are also adequately anticipated. If a simulator is funded and the researchers have limited experience using one, several years of lag typically occur before papers begin to appear at conferences. During this time, students and technicians figure out, often by trial and error, how to translate research questions into programming details, to develop complex experimental designs and to analyze vast quantities of data. As technical and experimental design experience is acquired, the scope of questions that can be addressed tends to be resolved. Thus, the actual fit of a simulator to research questions will be somewhat dependent on the understanding the researcher has when applying for funding and the eventual accumulation of research expertise using a simulator.

A mid-fidelity simulator is not needed for every research application. For instance, a very high-definition visual system could be connected to a gaming steering wheel, brake and accelerator. Such a high-low fidelity simulator might be used for visibility or sign comprehension research where vehicle response characteristics are not necessarily important measures. If a set of research questions can be framed with precise task requirements or fidelity, those physical components can be selected at higher levels of fidelity, whereas other simulator equipment not central to important tasks and measures is made to be nominally functional.

The fidelity of specific tasks that are the focus of a study often requires greater approximations of the real thing than a simulator can achieve. Simulators with fixed physical properties that cannot be appreciably modified have limited fidelity. In some cases, research can or cannot address a question based on inherent limitations. More likely, the researcher can address a question, but the results represent an approximation and may be open to criticism if peer reviewed. Questions that require specific lighting conditions (e.g., driving at dusk, dawn or night), high resolution of detail (e.g., complex signs), important response properties (e.g., electronic stability control) or tracking in the periphery (e.g., the approach of a high-speed train or merging in a work zone) may simply not be able to be addressed by a simulator with limited projection luminance, graphics engines, vehicle control models coupled to motion bases, and field of view.

How aspects of driving simulation with insufficient fidelity affect measures requires either comparison to on-road results or qualification with respect to the limitation. For instance, insufficient visibility of a sign due to resolution limitations will require that drivers be closer to the sign in order to read it. The overall time to view a sign would be expected to be less, which would also affect comprehension and legibility distance too. Comparison of simulator results to on-road testing is likely to reveal measures that are truncated in predictable ways as a function of the simulator limitation.

8) To what degree are the results from a simulator similar to those measured in the real world?

Probably the most often repeated yet also most difficult question about driving simulators is whether measures in a simulator mimic those measured when driving in the real world. This multi-level correspondence problem has broadly been called simulation validation, which has been defined as the replication of simulator and on-road tests to determine the extent to which measures correspond across contexts. Simulation validation has been a concern for at least 25 years (Blaauw, 1982). A number of important aspects of this topic are addressed by several chapters in this *Handbook*; namely, performance validity (see Mullen, Charlton, Devlin, & Bédard, this book, chap. 13) and cross-platform validation issues (see Jamson, this book, chap. 12). A number of authors have also reviewed simulation validation at different times (e.g., Blaauw, 1982; Godley, Triggs, & Fildes, 2002; Kaptein, Theeuwes, & Van der Horst, 1996).

A number of types of simulation validity are important. *Relative validity* indicates that the direction of change of a variable is in the same direction as a corresponding manipulation and measure in the real world (Kaptein et al., 1996). For example, if speed is measured at 5 km/h greater in a simulator than on-road across a range of speeds, the pattern of data would represent relative validity because it is in the same direction. *Absolute validity* is the extent that a manipulation in the real world when manipulated in the simulator produces the same or equivalent numerical change in the same measure (Blaauw, 1982, p. 474). For instance, for a specific condition, 50 km/h in the simulator for a given stretch of roadway is also measured at 50 km/h on-road.

Given the variability of driver performance, the exact correspondence of on-road and simulation measures is unlikely. Some authors suggest that finding measures that reflect absolute validity are not as important as finding measures that produce consistent changes in the dependent variable or relative validity (e.g., Kaptein et al., 1996; see in this book, chap. 3 by Kantowitz and chap. 13 by Mullen et al.; Törnos, 1998; Yan et al., 2008). Others have further argued that simulator companies should produce studies that establish validity for their products. For various reasons, the responsibility of validating simulators has fallen to researchers. Published studies on simulation validation are obviously a small fraction of operating simulators, especially if each simulator type represents a unique need to validate. A sizable body of gray literature on simulation validation appears in conference proceedings, technical reports and internal reports.

Sampling from published studies, simulation validation studies have generally been focused on findings of relative validity. For example, simulators have been found to have relative validity for speed (e.g., Blaauw, 1982; Törnos, 1998; Yan et al., 2008). Higher speeds were found in the tested simulators than in on-road segments. Lane-keeping measures were slightly more variable in the simulator compared to on-road measures in several studies (e.g., Blaauw, 1982; Reed & Green, 1999). However, the effects tend to vary depending on the simulator being evaluated and additional independent variables (e.g., experience, age, etc.). For example, higher speeds on-road were found in Godley et al. (2002) and lateral position in a tunnel shifted 13 cm more to the left in the actual tunnel (Törnos, 1998). Still others have used unique validation measures such as safety surrogates (Yan et al., 2008), which attempted to relate risk-taking maneuvers in a simulated intersection with the crash history of the actual intersection. Reed and Green (1999) measured many variables while drivers dialed a phone. The correlation between on-road and simulation variables ranged from 0.18 to 0.76. Thus, the use of some variables is consistent across simulation validation studies, whereas other variables are unique and cannot be compared across studies. The uniformity of relative validity is somewhat variable depending on the measure. Finally, compared to the number of variables represented in the dependent variable list of Table 5.2, relatively few have been used (or reported) in simulation validation studies, although this appears to be changing (see Mullen et al., this book, chap. 13).

The general comparison of variables in simulation validation can be abstractly illustrated. For a given measure (or aggregation of measures), a plot of obtained values from a simulator can be plotted against the values obtained for on-road performance under similar driving conditions. Thus, Figure 5.3 represents the correspondence of simulation and on-road measures. An exact correspondence of measures from road and simulator indicates that each is measuring exactly the same thing (i.e., absolute validity). Each line on the graph can be also thought of as the best fitting line of a scatterplot of data. The degree or extent that a given measure departs from an on-road variable is also plotted (relative validity). Several lines of relative validity are plotted (i.e., A & B). One illustrates a situation where on-road measures are consistently higher than those from simulation A. (The reverse is also plotted, B, where the measures are consistently lower.) Attribution as to why relative validity is higher or lower in either setting would then require systematic elimination of potential causes. For example, the finding that a simulator may yield higher lane variability may indicate that steering control is insufficiently tightly coupled. The situation becomes more complicated in circumstances depicted by line C of Figure 5.3, where some parameter yields higher values in on-road (versus simulator) measurement for a certain range of values (left-hand side of figure), yet lower values in another range (right-hand side). Whether line A is more valid than B and the extent to which line C is invalid are unresolved questions that require further empirical support and discussion.

Statistical tests are one way to determine if measures taken from on-road and simulation differ with respect to absolute validity or from other tests where relative validity is established, but

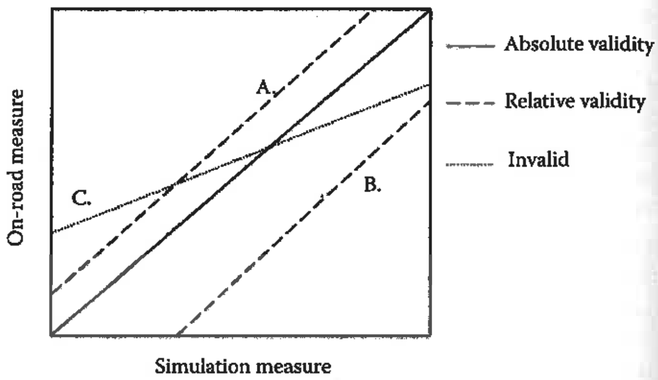

FIGURE 5.3 Absolute and relative validity expressed as a function of the correspondence of measures from on-road and simulator for identical contexts. See text for details.

what is the practical significance of a measure that differs more or less from absolute validity? Statistical testing of significant differences between simulation and on-road measures can be affected by unbalanced within-subjects treatment orders and strategies (e.g., boredom may lead to speed increases) (see Godley et al., 2002 for discussion). Measures are often collected over a limited range of values and are not necessarily treated as continuous variables. The relative validity of a simulator should be qualified relative to the range over which measures are collected. In addition, the relative validity of a collection of measures requires consideration.

9) How well do the results from simulator, on-road and more basic laboratory studies correspond with one another?

Researcher beliefs about the validity of driving simulation results range from advocacy to dismissive. Some scientists do not seem to be swayed by consistent evidence obtained from studies that compare on-road and simulator results that establish relative validity. In addition to direct comparisons within a study that determine simulation relative validity, measures can be compared across studies of on-road, simulator and laboratory. The correspondence of results across different experimental approaches is related to the previous question although comparisons are not restricted to simulation and on-road.

A number of recent studies have sought to statistically compare results arising from different methods. For instance, Caird, Willness, Steel, and Scialfa (2008) meta-analyzed studies from simulator, laboratory and on-road settings where the impact of cell phone conversations on driving performance were measured. Meta-analysis is a means to statistically compare the results of studies using effect sizes (Elvik, 2004; Rosenthal & DiMatteo, 2001). The categories of laboratory, simulator and on-road were defined. A laboratory study had at least an approximation of a tracking task ($N = 14$ studies). Driving simulator studies had a physical steering wheel coupled to an approximation of a visual traffic environment ($N = 18$). On-road studies were conducted on a test track or on ordinary roadways ($N = 5$). (Horrey & Wickens (2006) performed a similar analysis and found similar results but collapsed simulator and laboratory categories together.) For conversations while driving with hand-held or hands-free phones, on-road, simulator and laboratory settings all produced similar effects on reaction time and speed (i.e., slower reaction time and slightly slower speeds when conversing). As such, the belief that a single method—whether on-road, simulation, or laboratory—is best for approximating performance decrements does not seem to hold, at least for these tasks and measures. Approximating task demands (i.e., conversation) across settings produced similar effect sizes. The use of surrogate tasks to approximate actual performance is also addressed by Angell, this book, chap. 10.

Studies of in-vehicle information system interactions while driving have also compared results across laboratory, simulation and real-world settings (Santos, Meret, Mouta, Brookhuis, & de Waard, 2005). The abstract task that was created, and installed on a central screen, was a visual search task (called S-IVIS). The three research settings were able to detect, using lateral control measures, when drivers were looking down at the task, but simulation and field studies were also able to discriminate levels of search difficulty too. Comparisons across settings for other variables such as speed, speed variation, and distance headway were more complicated to interpret based on methodological differences across settings. The authors discuss a number of cautions about blanket generalizations.

As part of the same European project (called HASTE), Engström, Johansson and Östlund (2005) reported results of cognitive and visual tasks across real and simulated driving settings. The same visual search task was used (i.e., S-IVIS) and also an auditory memory task (i.e., ACMT), which was responded to by interacting with a touch screen at the top of the central column. Of interest, fixed (Volvo) and moving-based (VTI) simulators results for longitudinal, lateral, physiological, and eye movement measures were compared. The visual search task affected lateral control and reduced speed, whereas the cognitive memory task reduced lateral variability but not speed. Eye movements were concentrated on the center of the road when engaged in the cognitive task and to the search screen when engaged in the visual task. The pattern of measurement differences across settings provided similar results with some differences. Lateral variation was greater in the fixed simulator than the moving base simulator. In spite of this difference, replication of results across methods corresponds well when tasks are similar and method variance is minimal.

Validity and generalizability are both important experimental properties. The selection of a method that can provide an approximation of a generalizable effect of interest is fundamental to experimentation. Many years ago, Chapanis (1967) made the observation that laboratory tasks are rough approximations of real-life situations. Since then, human factors beliefs have wandered from accepting laboratory tasks as having external validity, which is the extent that results obtained from participants, environments and tasks necessarily relate to other participants, environments or tasks in real-life (Abelson, 1995). If task, environmental and population samples that are integrated into a study are similar to the same in real-life, generalizability is more likely to be achieved. Cost and time may dictate the methods or analytical approaches available to the researcher (Carsten & Brookhuis, 2005) and these decisions may limit the generalizability of a study. The danger in assuming that an effect in the laboratory or simulator holds to the same extent in the real world is that many other variables, not tested, may dilute or reduce the relative importance of a found effect (Chapanis, 1967). For example, many have wondered why more crashes do not occur when using cell phones when driving. One disturbing possibility, besides drivers not confessing to using a cell phone when a crash does occur, is that other conditions not tested may interact and mitigate the decrements of conversation on reaction time and speed in the real world.

Another related question concerns situations where data is available or are combined from different methodological approaches. What is the relative weight that should be given to driving simulation versus real-world results? The bias of some individuals is to place "real world" results on the highest platform of validity with other results relegated to lesser platforms. This naïve view is based on the intuitive belief that "real world" results, if obtained, have the highest validity. Consideration of the strengths and weaknesses of a given method—in essence, overcoming these potential biases—requires greater thought, scientific training and experience.

If a simulator study found similar results as a naturalistic study, which one is more valid? One finds a causal relationship between or among manipulation and measure(s), whereas the other describes an event or pattern of observations. In situations where the two approaches show convergent findings, the distinction between approaches may be less critical. However, if the results from a simulator conflicted with those found in a naturalistic study, which one purports the greater truth? Naturalistic studies are compelling because real behavior and crashes are captured and the videos can be seen, although it takes a while to capture these events. Driving simulation is relatively efficient at describing performance in a specific context and establishing causality among manipulations and measures. Observations made using the latter methodology represent a potential reality which may generalize to a limited set of conditions in the real world. That said, either method can yield low or high quality results (Nickerson, 1999). Judgment of research quality is an important expertise of researchers that can sort out how much weight or credibility to assign to a particular set of results (Wortman, 1994).

10) What are common threats to internal and external validity?

After reviewing hundreds of manuscripts, we have identified a number of recurrent threats to internal and external validity when conducting driving simulation studies (see Table 5.4). The purpose of listing the threats, solutions and references is to encourage researchers to take a careful pause before starting a study. Recognizing a design weakness or flaw before collecting data is preferable to realizing this prior to submitting a manuscript. Many threats to internal and external validity can also be learned from the school of "ego-crushing" manuscript reviews. Obviously reviewers could adopt a guiding or mentoring style when reviewing manuscripts that have been submitted.

These are high-level problem and solution descriptions and the details of implementation for a given study require further investigation and consideration by researchers. Issues of statistical analysis, reporting and interpretation are comprehensively addressed by others (see Boyle, this book, chap. 21; Cohen, 1990; see Dawson, this

TABLE 5.4 Common Threats to Internal and External Validity When Using Driving Simulation

Threat	Description	Solution	Reference
Failure to adequately screen participants.	Vision or health problems (among others) of certain individuals may contribute to experimental error that masks or distorts effects.	Test for outliers. Use appropriate tests to screen drivers. Use covariate analyses to remove offending variance.	Trick & Caird (this book, chap. 26); Rizzo (this book, chap. 46); Tabachnick & Fidell (2006)
Generalization issues.	Tasks, population samples and environments are not similar to whom or what you wish to generalize.	Qualify results according to generalizability limitations. Include similar tasks, samples, and environments to desired generalizations.	Abelson (1995); Kaptein et al. (1996)
Drop out due to simulator sickness.	Properties of the simulator or activities in the simulator cause participants to become sick.	Carefully screen at-risk participants. Reduce maneuvers that require a sweeping motion such as left or right turns (among many other remediation strategies).	Stoner et al. (this book, chap. 14); Trick & Caird (this book, chap. 26)
Non-randomization of participants, treatments or events.	Treatments, participants or events are not randomly assigned to levels of the independent variable. Events are predictably located within drives.	Randomize. Check for order effects.	Abelson (1995); Wilkinson et al. (1999)
Range or carry-over effects.	Multiple treatments are experienced by the same participant. The order of treatment and experience causes asymmetric effects.	Use between-subjects designs for different treatment levels. Limit the number of treatments and counterbalance. Check for order effects and qualify results accordingly.	Poulton (1982)
Low number of participants or observations per cell.	Few participants or observations per cell reduce the stability of results.	Run more participants or collect more observations. Conduct power analysis. When possible, consider more efficient experimental designs (e.g., within-subjects).	Cohen (1992); Wilkinson et al. (1999)
Visual fidelity distortions.	Incorrect luminance levels, missing or distorted monocular and binocular cues.	Qualify results according to visual limitations.	Andersen (this book, chap. 8); Wood & Chaparro (this book, chap. 28)
Control fidelity distortions.	The quality of steering, brake and acceleration is unlike that of actual vehicle performance.	Qualify results according to control limitations. Provide sufficient practice such that drivers meet some criterion level of performance or competence.	Flach et al., (this book, chap. 43); Greenberg & Blommer (this book, chap. 7)
Greater than 5% loss of data.	Equipment or subject problems cause a loss of data.	Use appropriate data substitution methods and report rationale. Drop participants, fix equipment and re-run. Use appropriate statistical tests if assumptions are violated.	Tabachnick & Fidell (2006); Siegal & Castellan (1988); Wilkinson et al. (1999)

book, chap. 22; Tabachnick & Fidell, 2006; Wilkinson et al., 1999). The final caveat of this list is that we have probably omitted a number of threats to internal and external validity so researchers should be vigilant about improving the quality of their experimentation.

11) Is there evidence that driving simulators have contributed to the advancement of knowledge since the 1970s?

Evans (2004, p. 190) makes the following argument:

> Driving simulators are far from new. A 1972 article refers to an earlier 1970 article listing 28 devices then in use, 17 of them in the U.S. Since the 1960s, driving simulators have incorporated moving bases and multiple movie projectors to provide visual information, including to the rear view mirror. The list was published in 1972. The research literature provides scant evidence that [the] research agenda was advanced by simulators, neither by those in existence in 1970, nor by the much larger number of far more expensive and sophisticated simulators that have been built.

To clarify, the list (or research agenda) that appears in Evans' book on page 191 includes traffic control devices, drug effects, the driver as a control element, vehicle characteristics, highway design, and driving conditions. To distill out a question, has driving simulation contributed to the development of these or other areas of knowledge since the 1970s?

In Figure 5.3, we plot the progression of studies that have used driving simulation as a method since 1970 in two relevant journals: *Accident, Analysis and Prevention (AAP) and Human*

Factors (HF). These journals were selected to get a general indication of international driving simulation research activity over a span of four decades. HF has been published by the Human Factors and Ergonomics Society since 1959 and is representative of activities in the U.S. and other countries. AAP was originally a European journal that has more recently (circa 2002) systematically expanded the number of studies published to take advantage of online dissemination. Both journals publish studies that focus on the driver and, to a lesser degree, on the underlying simulation technology.

Both journals were searched for studies that used driving simulation. Titles, abstracts and full papers were searched for the occurrence of <driving simulation> and other variants. Study inclusion criteria are modified from Caird et al. (2008) and Horrey and Wickens (2006) who, in their meta-analyses of cell phone studies, established categories for different methodological approaches; namely, laboratory, simulation and on-road. The simulation category included part-task, low-fidelity and high-fidelity simulators. In the present analysis, however, studies that fractured the coupling between perception and action are not included because a participant does not actively control anything. For instance, ratings or responses to picture or video presentations are not included.

How much research has been produced? As shown in Figure 5.4, the number of published studies using driving simulation in *AAP* has increased dramatically in the past four years. (The year 2008, which had 22 simulator studies, includes studies in press that may appear in 2009.) This rise seems to parallel the increase in the number of manuscripts accepted by *AAP*. Attribution of the increase to a growth in driving simulation research activity may be only partly correct. The increase probably reflects a modest rise in research activity, but also possible shifts in acceptance criteria at *AAP*. The increase in driving simulator studies in *HF* has been less dramatic over the entire 40-year span, though there is a recent upwards trend since the mid-90s.

For many years, research centers within various countries (e.g., TNO, NHTSA) developed driving simulators and productivity (i.e., peer-reviewed papers) was a reflection of the activity of these centers. Many technical difficulties in hardware and software had to be overcome to execute each research study. The current generation of driving simulators is significantly more flexible and allows researchers to develop and execute research on a much shorter time scale compared to previous generations of simulators. The cost of mid-fidelity simulation has declined as graphics and computation capabilities increased many fold in commonly available systems. The enhanced sophistication of simulators in this respect may also be an important contributor in the increase in published research using simulators.

What kinds of research have been produced? The purpose of Table 5.5 is to aggregate the general pattern of research activity in *AAP*. The subcategories roughly capture measures and/or manipulations of particular studies. A study was counted in several categories if multiple measures or manipulations captured the purpose of each study. For example, speed perception was counted under Speed and Perception/Attention. Clearly driving simulators have been used to evaluate devices for distraction, new road geometries, and traffic control as well as to understand lifespan individual differences among many other uses.

The list in Table 5.5 only partially resembles the list discussed by Evans (2004, p. 189), which indicates how research simulators were being used in 1972. To be fair, many of the studies that appear in Table 5.5 have been published since the original criticism by Evans. Finding research gaps is as simple as combining categories or factors until that research has not been performed previously. Some questions or combinations are obvious, whereas others are deeper and transcend independent variable combinations.

There are some limitations to simply counting and classifying studies from journals. Many driving simulation studies are conducted for a sponsoring agency (e.g., NHTSA). The results

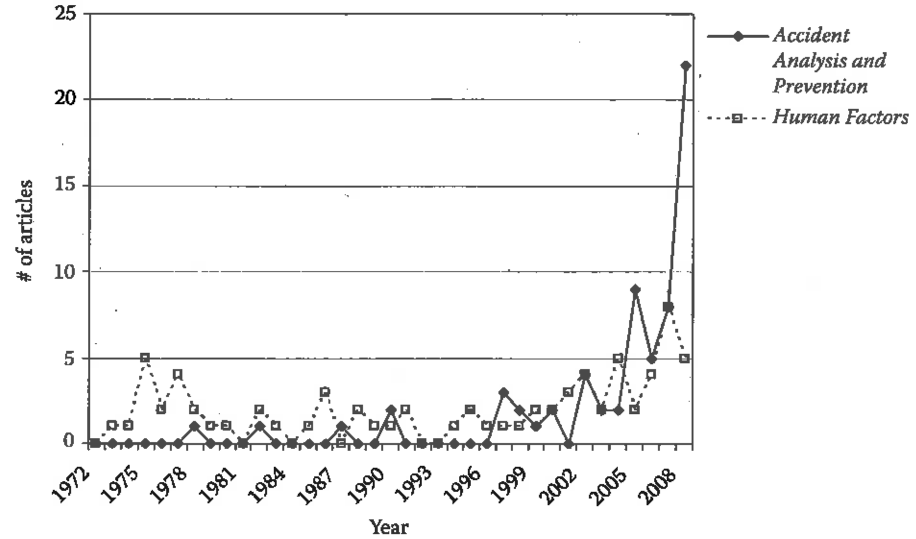

FIGURE 5.4 The number of simulator studies by year published in *Accident, Analysis and Prevention* and *Human Factors* from 1970 to 2008. (Not plotted *HF*: 1964 = 1, 1968 = 1.)

TABLE 5.5 Types and Subcategories of Driving Simulation Studies Published in *Accident, Analysis and Prevention* From 1969 to 2008

Study Type	Subcategories
Alcohol (4) or Drugs (1)	On-board detection (1), Age (1), Skilled drivers (1), Other (2)
Distraction (11)	Cell phone (5), Conversation (1), Feedback (1), Eating (1), Multiple sources (2), MP3 (1)
Devices (2)	Warnings (1), Audio Speakers (1)
Fatigue or Sleepiness (10)	Napping (1), Other intervention (1), Simple versus complex measures (1), Sleep apnea (2), Physiological measures (1), Other (4)
General Individual Differences (8)	Expert/Novice (1), ADHD (1), Gender (2), Police (1), Truck Drivers (1), Teens (1), Training (1)
Older Drivers (8)	Self ratings (1), Individual differences testing (4), Pedal errors (1), Other (2)
Pedestrians (4)	Unfamiliar traffic direction (1), Age/development (2), Timing (1)
Perception/Attention (7)	Visibility (3), Fog (2), Speed selection (1), Visual fields (1)
Roadway (12)	Curves (2), Rural two-lane (1), Intersections (3), Horizontal curves (1), Tunnels (2), Width (2) Traffic lights (1), Lane treatments (1)
Simulation Validation (4)	Intersections (1), Tunnel (1), Speed (1), Older drivers (1)
Speed (5)	Two-lane rural road (1), Perception (1), Gender (1), Fog (1), Curves (1)

are put into a technical report and may or may not be published in a journal. For instance, Meister (1995) provides pointers to many of these flight and driving technical reports that appeared from the 1950s to the 1980s and reference backtracking reveals many others. Technical reports often have considerably more details and analyses than condensed peer-reviewed publications, including simulator and methods descriptions. Thus, reliance on journals such as *AAP* and *HF* probably underestimates the true scope and overall activity that driving simulation has contributed to knowledge. Moreover, some editors may choose to exclude certain methods if they feel the results are not valid.

The contribution of *AAP* and *HF* to knowledge advancement based on published driving simulation studies is open to criticism. At the time that a study is published, reviewers and editors may not be aware whether or not a particular study advances the state of knowledge over that which existed prior to publication. Awareness of research worldwide is daunting but facilitated by search engines, citation indexes and custom publication alerts. Determination of the relative contribution of a given study to the collective body of knowledge is fallible and the relative influence or impact of a given study varies. The collective contribution of the body of research across time does not simply reduce to a journal's impact factor multiplied times the number of studies on driving simulation that appear in it (among other metrics) (e.g., Addair & Vohra, 2003). The qualitative dimension of influence on subsequent thought, method, policy and countermeasure is not captured by quantitative reductions. Suffice it to say that the collective impact of these papers has been constant and is now accelerating.

12) Based on the progress of software, hardware and projection capabilities, where will the future of driving simulation lead?

While this chapter was being written, U.S. and Canadian government policy has decided to bail out the Detroit 3 automakers as a long-term recession progressed worldwide. A quarter previously, worldwide vehicle manufacturing was looking towards India's Nano by Tata, a low-cost vehicle, and China's increasing demand for automobiles for the middle class. Globally this outlook has completely changed. Without question the current financial oscillations will affect budgets for research and development within corporations, local and national governments. In turn, the use of driving simulation to address a range of research and development agendas will likely be curtailed especially as large-scale commitments to naturalistic studies have been made in the U.S. and Europe.

Beyond the uncertainty of the current global recession (or depression), future directions will resemble past and current technological vectors. First, computer processor speed, in both graphics and operations, will increase in accord with Moore's Law, which predicts the cost of processing power halving every 18 months. However, the translation of this speed into graphics engines and usable software for driving simulators is imperfect compared to the same translation into computer games. An off-the-shelf gaming system running a driving game looks significantly better than the graphics of most moderate-fidelity driving simulators. Although sitting in an actual car surrounded by screens is compelling, simulation graphics capabilities are impoverished compared to what can be played at home. The reason for this obvious discrepancy is the close relationship that has evolved between game developers and graphics chip producers. Driving simulator companies are not part of this relationship.

Second, the usability of most driving simulator software frequently requires a computer programmer to overcome numerous idiosyncratic interactions, scripting and programming requirements to execute important functions. This becomes a perennial cost center in addition to software licensing and hardware upgrades. In contrast, easy-to-use experimenter interfaces provide a means of addressing more research questions by a wider range of students and researchers. A large visual database and an easy-to-use interface probably facilitate the production of more research over the half-life of a simulator. The limited marketplace for driving simulators prevents small, undercapitalized companies from adequately designing software, visual databases, graphics engines and user interfaces among other aspects of driving simulators. Market forces such as the adoption of training curricula or driver testing using driving simulation may create a larger market for driving simulators, but these simulators may not necessarily be of a flexible variety because curriculum modules will become part of the sale.

Third, increases in storage media size will permit researchers to collect more variables at higher sampling rates, including multiple digital video streams, all at lower price points. However, what to do with all the data will remain an issue. Reduction of data and the conversion of it into useful and insightful results will require that researchers know what to look for, where to find it and how to convey it coherently. Some argue that only theory can guide researchers on where to look. If only it was so simple (Meehl, 1993). Only a bricolage of observations and theories will sufficiently describe and explain driver performance and behavior across contexts (e.g., Dewar & Olson, 2007; Groeger, 2000; Shinar, 2007). Current and future generations of simulation may allow for greater insight into more and more subtle combinations of relationships. For instance, dynamic visualizations of complex variable relationships and re-creations of actual crashes (e.g., National Transportation Safety Board, 2003) are important directions that need to be pursued.

Fourth, research with driving simulators should make a difference and address important gaps in knowledge (e.g., Pedean et al., 2004; Sivak et al., 2007). In particular, understanding why and how drivers are killed and injured is an important endeavor (Evans, 2004) as is determining if countermeasures are effective (Elvik, 2004). Identifying where and when the greatest number of drivers are killed or injured is a form of research triage that can prioritize a research agenda. Current hot topics in research, for instance driver distraction, can influence researchers away from considering important crash contributors that are not constantly in the media. In addition, the business of attracting contract or grant funding to feed simulator operational and soft-funding budgets can obscure the importance of pursing traffic safety questions which do not necessarily fit into funding research agendas of the day.

Finally, will researchers in the future look back at the technology of driving simulators today with the same detached curiosity we have when examining Figures 5.1 and 5.2? Will they exclaim, "How did they do what they did with such primitive technology?" Hopefully, "Why did they do that (kind of research)?" will be less of a concern. Will driving simulation be indistinguishable from real driving (see Hancock & Sheridan, this book, chap. 4)? Will our children's children still drive at all (or swim)?

Key Points

- Studies that have used driving simulation since the 1970s have contributed to a number of important research areas including human factors and traffic safety.
- Driving simulators are inherently neither bad nor good methodologically. How simulators are used by researchers can result in high-quality research or results that are open to criticism.
- Threats to internal and external validity in driving simulation can, in part, be addressed by attention to a number of methodological details.
- Driving performance in simulators may provide an optimistic view of drivers' capabilities relative to actual behavior.
- In the future, the cost of driving simulators will decrease, realism will increase, and applications are likely to expand.

Keywords: Epistemology, History, Philosophy of Science, Simulation Validation

Acknowledgments

Don Fisher, Mary Lesch, Marvin Dainhoff, Andrew Mayer, Shaunna Milloy, Pierro Hirsh, Lana Trick, & Matt Rizzo provided helpful guidance, comments and edits to early versions of this chapter. The Canadian Foundation for Innovation (CFI) funded the University of Calgary Driving Simulator (UCDS) and the AUTO21 Networks of Centers of Excellence (NCE) funded numerous research studies using it.

Glossary

Absolute validity: The extent to which a manipulation of a variable in the real world produces the same or equivalent change in the same measure when manipulated in a driving simulator (Blaauw, 1982, p. 474).

Comprehensiveness: The extent to which operational functions and environmental characteristics, etc., are reproduced in a simulator (Meister, 1995, p. 204).

External validity: The extent to which results obtained from participants, environments and tasks obtained in one setting necessarily relate to other participants, environments or tasks in other, broader settings (Abelson, 1995).

Fidelity or realism: The degree to which reality is matched in a simulation (Kantowitz, this book, chap. 3).

Interestingness: To change what people believe about an important topic (Abelson, 1995, p. 13).

Internal validity: The conclusion that a manipulation of an independent variable results in a corresponding effect on a dependent variable and does not result from some other known or unknown influence (Abelson, 1995).

Physical fidelity: The extent to which a physical variable in a simulator (e.g., roll) corresponds to its operationally equivalent component in the real world (Lee, 2005, p. 88; see also, Greenberg & Blommer, this book, chap. 7).

Psychological fidelity: The degree to which the simulation task is perceived by participants as being a duplicate to the operational task (Meister, 1995, p. 206).

Relative validity: The direction of change of a variable is in the same direction as a corresponding manipulation and measure in the real world (Kaptein et al., 1996).

Simulation validation: The replication of simulator and on-road tests to determine the extent to which measures correspond across contexts.

Web Resources

The *Handbook's* web site contains supplemental materials for this chapter including two color figures:

Web Figure 5.1: Link flight simulator (left) and terrain map (right) displayed at the Smithsonian Air and Space Museum in Washington, D.C.

Web Figure 5.2: Link flight simulator comic showing simulator sickness or vertigo circa 1942 displayed at the Canadian Air Force Hall of Fame Museum in Wetaskiwin, Alberta.

Key Readings

Abelson, R. P. (1995). *Statistics as principled argument.* Hillsdale, NJ: Lawrence Erlbaum.

Evans, L. (2004). *Traffic safety.* Bloomfield Hills, MI: Science Serving Society.

Gibson, J. J., & Crooks, L. E. (1938). A theoretical field-analysis of automobile driving. *American Journal of Psychology, 51*(3), 453–471.

Shinar, D. (2007). *Traffic safety and human behavior.* Amsterdam: Elsevier.

References

Abelson, R. P. (1995). *Statistics as principled argument.* Hillsdale, NJ: Lawrence Erlbaum.

Addair, J. G., & Vohra, N. (2003). The explosion of knowledge, references and citations: Psychology's unique response to a crisis. *American Psychologist, 58,* 15–23.

Blaauw, G. J. (1982). Driving experiments and task demands in simulator and instrumented car. *Human Factors, 24,* 473–486.

Boer, E. R. (2001). Behavioral entropy as a measure of driving performance. *Proceedings of the first international driving symposium on human factors in driving assessment, training and vehicle design* (pp. 225–229). Aspen, CO.

Caird, J. K., Chisholm, S., Edwards, C., & Creaser, J. (2007). The effect of yellow light onset time on older and younger drivers' perception response time (PRT) and intersection behavior. *Transportation Research, Part F: Traffic Psychology and Behavior, 10*(5), 383–396.

Caird, J. K., Chisholm, S., & Lochhart, J. (2008). The effect of in-vehicle advanced signs on older and younger drivers' intersection performance. *International Journal of Human Computer Studies, 66*(3), 132–144.

Caird, J. K., & Dewar, R. E. (2007). Driver distraction. In R. E. Dewar & P. L. Olson (Eds.), *Human factors in traffic safety* (2nd ed., pp. 195–229). Tucson, AZ: Lawyers & Judges Publishing.

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

Carsten, O., & Brookhuis, K. (2005). The relationship between distraction and driving performance: Towards a test regime for in-vehicle information systems. *Transportation Research, Part F: Traffic Psychology and Behavior, 8*(2), 75–77.

Chapanis, A. (1967). The relevance of laboratory studies to practical situations. *Ergonomics, 10*(5), 557–577.

Chisholm, S. L., Caird, J. K., & Lockhart, J. (2008). The effects of practice with MP3 players on driving performance. *Accident Analysis & Prevention, 40,* 704–713.

Cohen, J. (1990). What I have learned (so far). *American Psychologist, 45*(12), 1304–1312.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159.

De Silva, H. R. (1936). On an investigation of driving skill. *The Human Factor, 10,* 1–13.

Dewar, R., & Olson, P. (2007). *Human factors and traffic safety* (2nd ed.). Tucson, AZ: Lawyers and Judges.

Dingus, T., Klauer, S. G., Neale, V. L., Petersen, A., Lee, S. E., Sudweeks, J., et al. (2006). *The 100-car naturalistic driving study, Phase II—Results of the 100-car field experiment* (Rep. No. DOT HS 810 593). Washington, D.C.: National Highway Traffic Safety Administration.

Elvik, R., & Vaa, T. (2004). *The handbook of road safety measures.* Amsterdam, the Netherlands: Elsevier Science.

Endsley, M. R. (2000). Direct measurement of situation awareness: Validity and use of SAGAT. In M. R. Endsley & D. J. Garland (Eds.), *Situation awareness and measurement.* Mahwah, NJ: Lawrence Erlbaum.

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

Evans, L. (1991). *Traffic safety and the driver.* New York: Van Nostrand Reinhold.

Evans, L. (2004). *Traffic safety.* Bloomfield Hills, MI: Science Serving Society.

Fisher, P., Kubitzki, J., Guter, S., & Frey, D. (2007). Virtual driving and risk-taking: Do racing games increase risk-taking cognitions, affect, and behaviors? *Journal of Experimental Psychology: Applied, 13*(1), 22–31. .

Gaba, D. (2004). The future vision of simulation in health care. *Quality Safety in Health Care, 13,* 2–10.

Gibson, J. J. (1986). *The ecological approach to visual perception.* Hillsdale, NJ: Lawrence Erlbaum.

Gibson, J. J., & Crooks, L. E. (1938). A theoretical field-analysis of automobile driving. *American Journal of Psychology, 51*(3), 453–471.

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

Godthelp, H., Milgram, P., & Blaauw, G. L. (1984). The development of a time-related measure to describe driving strategy. *Human Factors, 26*(3), 257–268.

Green, M. (2000). "How long does it take to stop?" Methodological analysis of driver perception-brake times. *Transportation Human Factors, 2*(3), 195–216.

Green, P (2007). Where do drivers look while driving (and for how long)? In: R. E. Dewar & R. Olsen (Eds.), *Human Factors in Traffic Safety* (2nd ed., pp. 57–82). Tucson, AZ: Lawyers & Judges Publishing.

Green, P. (2008). Developing complex crash warning simulations for human factors evaluations. *Proceedings of the human factors and ergonomics society 52nd annual meeting* (pp. 1865–1869). Santa Monica, CA: Human Factors and Ergonomics Society.

Green, P., Cullinane, B., Zylstra, B., & Smith, D. (2004). *Typical values for driving performance with emphasis on the standard deviation of lane position: A summary of literature* (Tech. Rep. SAVE-IT, Task 3a). Ann Arbor, MI: University of Michigan Transportation Research Institute (UMTRI).

Groeger, J. A. (2000). *Understanding driving: Applying cognitive psychology to a complex everyday task.* Philadelphia, PA: Taylor & Francis.

Hart, S. G., & Staveland, L. E. (1988). Development of the NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139–183). Amsterdam, the Netherlands: North Holland.

Human Machine Interface and the Safety of Traffic in Europe (HASTE). (n.d.). *Internal deliverable: WP2 Pilot Dependent Variables.*

Horrey, W. J., & Lesch, M. F. (2009). Driver-initiated distractions: Examining strategic adaptation for in-vehicle task initiation. *Accident Analysis & Prevention, 41*, 115–122.

Horrey, W. J., & Simons, D. J. (2007). Examining cognitive interference and adaptive safety behaviors in tactical vehicle control. *Ergonomics, 50*(8), 1340–1350.

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

Horrey, W. J., & Wickens, C. D. (2007). In-vehicle glance duration: Distributions, tails, and a model of crash risk. *Transportation Research Record, 2018*, 22–28.

Horrey, W. J., Wickens, C. D., & Consalus, K. P. (2006). Modeling drivers' visual attention allocation while interacting with in-vehicle technologies. *Journal of Experimental Psychology: Applied, 12*(2), 67–78.

International Standards Organization. (2002). *Road vehicles—Measurement of driver visual behavior with respect to transport information and control systems—Part 1: Definitions and parameters,* ISO committee Standard 15007-1, Geneva, Switzerland: International Standard Organization.

Jones, D. G., & Kaber, D. B. (2005). Situation awareness measurement and the situation awareness global assessment technique. In N. Stanton, A. Hedge, K. Brookhuis, E. Salas, & H. Hendrick (Eds.), *Handbook of human factors and ergonomics methods* (pp. 42-1–42-8). New York: CRC Press.

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

Lamble, D., Kauranen, T., Laakso, M., & Summala, H. (1999). Cognitive load and detection thresholds in car following situations: Safety implications for using mobile (cellular) telephones while driving. *Accident Analysis & Prevention, 31*, 617–623.

Lee, A. T. (2005). *Flight simulation.* Aldershot, United Kingdom: Ashgate.

Lee, J. D., & Strayer, D. L. (2004). Preface to the special section on driver distraction. *Human Factors, 46*(4), 583–586.

Meehl, P. (1993). Theoretical risks and tabular risks: Sir Karl, Sir Ronald and the slow progress of soft psychology. In C. A. Anderson & K. Gunderson (Eds.), *Selected philosophical and methodological papers of Paul E. Meehl* (pp. 1–42). Minneapolis, MN: University of Minnesota Press.

Meister, D. (1995). Simulation and modeling. In J. R. Wilson & E. N. Corlett (Eds.), *Evaluation of human work* (2nd ed., pp. 202–228). London: Taylor and Francis.

Miles, G. H., & Vincent, D. F. (1934). The Institute's tests for motor drivers. *The Human Factor, VIII* (7–8), 245–257.

Mulder, L. J. M., de Waard, D., & Brookhuis, K. A. (2005). Estimating mental effort using heart rate and heart rate variability. In N. Stanton, A. Hedge, K. Brookhuis, E. Salas, & H. Hendrick (Eds.), *Handbook of human factors and ergonomics methods* (pp. 20-1–20-8). Boca Raton: CRC Press.

National Transportation Safety Board. (2003). *Ford Explorer Sport collision with Ford Windstar Minivan and Jeep Grand Cherokee on Interstate 95/495 near Largo, Maryland, February 1, 2002* (Rep. No. NTSB/HAR-03/02). Washington, DC: National Transportation Safety Board.

Nickerson, R. S. (1999). Basic and applied research. In R. J. Sternberg (Ed.), *The nature of cognition* (pp. 409–444). Cambridge, MA: MIT Press.

Olson, P. L., & Farber, E. (2003). *Forensic aspects of driver perception and response* (2nd ed.). Tucson, AZ: Lawyers and Judges, Inc.

Olson, P. L., & Sivak, M. (1986). Perception-response time to unexpected roadway hazards. *Human Factors, 28*, 91–96.

Pauzié, A. (2008). A method to assess the driver mental workload: The driving activity load index (DALI), *IET Intelligent Transport Systems, 2*(4), 315–322.

Peden, M., Scurfield, R., Sleet, D., Mohan, D., Hyder, A. A., Jarawan, E., & Methers, C. (2004). *World report on road traffic injury prevention.* Geneva: World Health Organization.

Poulton, E. C. (1982). Influential companions: Effects of one strategy on another in the within-subjects designs of cognitive psychology. *Psychological Bulletin, 91*(3), 673–690.

Reed, M. P., & Green, P. A. (1999). Comparison of driving performance on-road and in a low-cost simulator using a concurrent telephone dialing task. *Ergonomics, 42*(8), 1015–1037.

Rosenthal, R., & DiMatteo, M. R. (2001). Meta-analysis: Recent developments in quantitative methods for literature reviews. *Annual Review of Psychology, 52*, 59–82.

Santos, J., Meret, 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. *Transportation Research, Part F: Traffic Psychology and Behavior, 8*, 135–146.

Schiff, W., & Arnone, W. (1995). Perceiving and driving: Where parallel roads meet. In Hancock, P. A., Flach, J., Caird, J. K., & Vicente, K. (Eds.), *Local applications of the ecological approach to human machine systems* (pp. 1–35). Hillsdale, NJ: Lawrence Erlbaum.

Sheridan, T. B. (1970). Big brother as driver: New demands and problems for the man at the wheel. *Human Factors, 12*(1), 95–101.

Shinar, D. (2007). *Traffic safety and human behavior.* Amsterdam: Elsevier.

Siegal, S., & Castellan, N. J. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). New York: McGraw-Hill.

Sivak, M., Luoma, J., Flannagan, M. J., Bingham, C. R., Eby, D. W., & Shope, J. T. (2007). Traffic safety in the U.S.: Re-examining major opportunities. *Journal of Safety Research, 38*, 337–355.

Society of Automotive Engineering. (2000). *Definition and measures related to the measurement of driver behavior using video based techniques* (SAE Recommended Practice J2396). Warrendale, PA: Society of Automotive Engineering.

Strayer, D. L., & Drews, F. A. (2004). Profiles in driver distraction: Effects of cell phone conversations on younger and older drivers. *Human Factors 46*(4), 640–649.

Stutts, J. C., Faeganes, J., Reinfurt, D., Rodgman, E., Hamlett, C., Gish, K., et al. (2005). Drivers' exposures to distractions in their natural environment. *Accident Analysis & Prevention, 37*(3), 1093–1101.

Summala, H. (2000). Brake reaction times and driver behavior analysis. *Transportation Human Factors, 2*(3), 217-226.

Tabachnick, B. G., & Fidell, L. S. (2006). *Using multivariate statistics* (5th ed.). Boston, MA: Pearson, Allyn, & Bacon.

Taylor, J. L., Kennedy, Q., Noda, A., & Yesavage, J. A. (2007). Pilot age and expertise predict simulator performance: A 3-year longitudinal study. *Neurology, 68*, 648–654.

Tijerina, L., Barikman, F. S., & Mazzae, E. N. (2004). *Driver eye glance behavior during car following* (Rep. DOT HS 809 723). Washington, D.C.: National Highway Traffic Safety Administration.

Tijerina, L., Johnson, S., Parmer, E., & Winterbottom, M. D., & Goodman, M. (2000). *Driver distraction with wireless telecommunications and route guidance systems* (Rep. No. DOT HS 809-069). Washington, D.C.: National Highway Traffic Safety Administration.

Törnos, J. (1998). Driving behavior in a real and a simulated road tunnel—A validation study. *Accident, Analysis and Prevention, 30*, 497–503.

Tufano, D. R. (1997). Automotive HUDs: The overlooked safety issues. *Human Factors, 39*(2), 303–311.

van Winsum, W., Brookhuis, K. A., & de Waard, D. (2000). A comparison of different ways to approximate time-to-lane crossing (TLC) during car driving. *Accident Analysis & Prevention, 32*(1), 47–56.

Wickens, C. D. (2001). *Attention to safety and the psychology of surprise.* Proceedings of the 2001 Symposium on Aviation Psychology. Columbus, OH: Ohio State University.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomic Science, 3*(2), 159–177.

Wickens, C. D., Todd, S., & Seidler, K. (1989). *Three-dimensional displays: Perception, implementation and applications* (Tech Rep. CSERIAC SOAR 89–001). Wright-Patterson Air Force Base, OH: US Air Force, Armstrong Aerospace Medical Research Laboratory (AFAAMRL).

Wierwille, W. W., & Fung, P. P. (1975). Comparison of computer generated and simulated motion picture displays in a driving simulation. *Human Factors, 17*(6), 577–590.

Wilkinson, L., & Task Force on Statistical Inference (1999) Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*(8), 594–604.

Wortman, P. M. (1994). Judging research quality. In: H. Cooper & L. V. Hedges (Eds.), *Handbook of research synthesis* (pp. 97–109). New York: Russell Sage Foundation.

Yan, X., Adbel-Aty, M., Radwan, E., Wang, X., & Chilakapati, P. (2008). Validating a driving simulator using surrogate safety measures. *Accident Analysis & Prevention, 40*, 274–288.

Young, M. S., & Stanton, N. A. (2005). Mental workload. In N. Stanton, A. Hedge, K. Brookhuis, E. Salas, & H. Hendrick (Eds.), *Handbook of human factors and ergonomics methods* (pp. 39-1–39-9). Boca Raton: CRC Press.

Zijlstra, F. R. H. (1993). *Efficiency in work behavior. A design approach for modern tools.* Doctoral dissertation, Delft University of Technology. Delft, the Netherlands: Delft University Press.