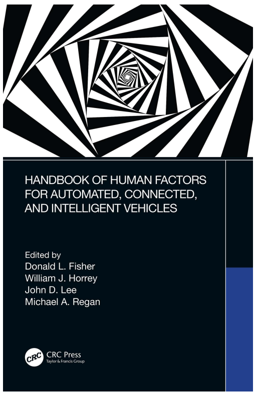

Handbook of Human Factors for Automated, Connected, and Intelligent Vehicles

Donald L. Fisher, William J. Horrey, John D. Lee, Michael A. Regan

HF Considerations When Testing and Evaluating ACIVs

Publication details
https://test.routledgehandbooks.com/doi/10.1201/b21974-22
Sheldon Russell, Kevin Grove

22 HF Considerations When Testing and Evaluating ACIVs

Sheldon Russell and Kevin Grove
Virginia Tech Transportation Institute Center
for Automated Vehicle Systems

CONTENTS

KEY POINTS

- Accurately characterizing the automated driving system(s) present in a platform is critical for testing, allowing participant training materials to be developed that accurately inform participants about system function, and aids in the selection of testing scenarios throughout the development process.
- Testing of commercial vehicles involves unique operational domains as well and specialized automation such as platooning.
- Methods for testing in early development are expected to be iterative, and to build a foundation for later stages of prototype testing.
- Analysis of existing naturalistic driving data will allow testing of models of automated features to undergo bench testing, and also aid in selection of testing scenarios.
- Simulator testing provides a method of iterative testing of early feature designs with a large degree of experimental control, but reduced external validity.
- Mid-development testing approaches should include on-road testing, using a prototype or Wizard of Oz approach to increase external validity of testing, at the expense of iterative testing.
- Late stage testing (including post-development) can include naturalistic driving studies in which large datasets of drivers actively using automation are collected.

22.1 INTRODUCTION

The purpose of this chapter is to provide an overview for feature development and human subjects' testing of various aspects of automated driving systems. Considerations for driving automation when testing heavy commercial vehicles (e.g., tractor trailers, busses) are also included. Although they can exist as stand-alone systems, connected vehicle features are not covered separately in this chapter; for the purposes of testing they are described here as a feature of an automated driving system. Other Level 1 features (e.g., automated emergency braking, blind spot warning, etc.) may also be part of the driving system, but are referred to generally here as Advanced Driver Assistance Systems (ADAS).

The terminology used in this chapter is intended to be generally consistent with SAE J3016 in referring to the driving automation system and/or features of said system (rather than a vehicle) (SAE International, 2016). However, some distinctions in classification are noted. For the purposes of this chapter driving automation is considered to be any sustained automation of both lateral and longitudinal functions (i.e., Level 2 and above). Furthermore, SAE J3016 defines a specific type of alert unique to automated driving systems, a Request to Intervene (RTI). An RTI is an alert or notification from the automation system to the driver that an intervention is needed. Per SAE J3016, RTIs are defined in the context of Level 3 (or above) systems; however, one could consider "hands-on wheel type alerts" designed to keep the driver engaged in the driving task as RTIs as well (insofar as an intervention is requested from the driver; Russell et al., 2018).

Previous chapters have covered issues that are fundamental to the human driver, driving automation, driving in general, as well as potential problems associated with driving automation and connected vehicle systems. Potential solutions to these problems such as driver training (this Handbook, Chapter 18), driver monitoring (this Handbook, Chapter 11), and human–machine interface (HMI) designs have also been presented (this Handbook, Chapters 15, 16). No matter how principled, any potential solution is just that until it is tested and evaluated. This chapter will include high-level overviews of steps critical to testing of driving automation systems. These steps include system characterization, commercial vehicle considerations, and testing methodologies (e.g., driving simulator, test track or live road experiments, and naturalistic driving).

SAE standard J3018 provides some guidance for testing at automation Levels 3 and above (SAE International, 2015a). The standard advocates a graduated approach where the expertise required for testing scenarios decreases over the course of testing, while the complexity of the testing scenario increases over the course of development. The standard provides definitions for expert test drivers, experienced test drivers, and novice test drivers. Expert test drivers are typically engineers who are designing the automated features themselves and can interact with the systems at the mechanical and software level; experienced drivers are trained on the systems but are unable to interact at the software level; and finally novice drivers have received only cursory training (if any). Testing locations and road scenarios should also be graduated in complexity; potential testing variables listed in the standard include

- Location
    - Test track
    - Closed campus operations (e.g., military base, corporate or university campus)
    - Public roads
- Roadway Type
    - Limited access freeway
    - Highway (single or multi-lane),
    - Arterial roads
    - Residential streets
    - Driveway, parking lot, or structure
- Traffic Environment
    - Traffic density
    - Vehicles
    - Pedestrians
    - Signage
    - Irregular—construction, crash scenes, road detours, flooding
    - Complex intersections, merges
    - Regional variations in road design
    - Traffic control devices (signals, signs, curbs, guardrails, etc.)
- Time of Day
    - Lighting conditions (day vs. night)
- Seasonal
    - Weather conditions

While providing some guidance, SAE J3018 does not provide any specific methodology for testing during phases. Furthermore, a system and associated features will undoubtedly go through many different tests throughout their development cycle. Assuming that a system is being developed from start to finish, the methodologies described herein are intended to build upon one another across the development cycle (e.g., data modeling in early development leads to scenarios for driving simulator testing, production systems are then tested via naturalistic driving studies (NDSs)). The scope of graduated testing is broad; it includes tests that may not consider the driver (i.e., engineering evaluations). While these tests are critical to the development cycle, the scope of this chapter is focused toward human subjects testing. It is assumed that engineering evaluations of the systems and components have been completed, and the features themselves are operable.

Testing is not limited to systems early in the development cycle. Testing of post-development cycle (i.e., commercially available) driving automation will always be necessary. Post-development testing can be categorized into two primary focuses of post-release monitoring and testing of broader safety benefits. Post-release monitoring refers to maintaining system reliability and performance once features are out in the world, with a focus on "black box" or vehicle data monitoring by an original equipment manufacturer (OEM) or other researcher. Analysis of this data may lead to over-the-air software updates (e.g., Tesla autopilot software changes) or otherwise inform future system designs. Furthermore, automated systems that have been deployed on public roads may allow for testing safety benefits within the larger transportation system, for example, crash rate comparisons between vehicles equipped with driving automation and non-equipped vehicles, which may then lead to rulemaking and/or policy considerations. Although distinct, these two approaches are complimentary. Post-release monitoring and safety benefit testing results may lead to insights into the overall capabilities and limitations at each level of automation and understanding limitations of current platforms should inform the design of future iterations of driving automation. Figure 22.1 provides a summary of testing throughout the development cycle, beginning with heuristic evaluation of the concept, usability testing of a prototype, user studies of pre-production models, and in-service monitoring of the released product.

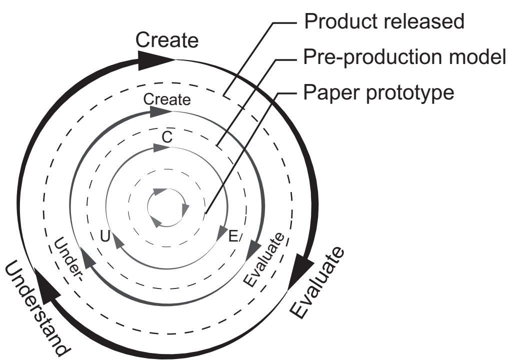

FIGURE 22.1 Overview of the development cycle. (From Lee, Wickens, Liu, & Boyle, 2017.)

22.1.1 Testing Approach: Avoiding Significant But Not Meaningful

Avoiding non-meaningful results is an unstated goal that is key to scientific discovery in general. Still it bears repeating, that when testing driving automation systems, it is critical that the researchers make their best effort to avoid focusing on statistical significance at the expense of practical significance, even if the result is a non-significant statistical test. The testing methodology, the driving scenario tested, and the participant characteristics will all determine the nature of results obtained and will all be critical for validity of testing. Additionally, there is a clear distinction in testing approaches; attempting to troubleshoot or find failure modes will lead to a different testing procedure than testing theory using inferential statistics (Lee et al., 2017). It will be up to the researchers designing and conducting the tests to select an appropriate series of tests as part of feature development and/or evaluation. Not an easy task to be sure. Methodologies should be layered, and testing scenarios should be designed that cover a broad spectrum of situations and use cases. Scenarios may be orchestrated in order to test a known limitation or edge case of the system. Alternatively, scenarios may be exploratory, in that they are designed to uncover edge cases rather than test them. The following sections include discussions of the basic underpinnings of system characterization and participant training which are relevant to nearly any test methodology.

22.1.2 Driving Automation Characterization

At the risk of stating the obvious, it is of utmost importance before any tests are designed that the intended function of the driving automation system is clearly understood by the researchers, a process referred to here as characterization. Whether the system of interest is in the conceptual phase or already in production, an understanding of the system functions should be developed by the researcher(s). Characterization supports multiple aspects of testing and evaluation, at multiple phases of development. In particular, it serves as a heuristic evaluation of a system, helps guide research question development and testing procedures, serves as a foundation for training participants on system function, and finally guides data sampling and reduction in naturalistic studies. Characterization includes identifying the expected performance envelope(s) (e.g., speed ranges, steering torque, etc.), the timing, modality, and type(s) of the RTIs, the presence of ADAS (e.g., collision mitigation braking), alerts associated with ADAS features (e.g., forward collision warning—FCW, blind spot warning), and so on. Characterization needs will vary and evolve depending on the stages of development and testing (e.g., early, mid, late) as capabilities are refined.

Any particular make, model, or brand of feature will have qualities that will alter driver interactions with the systems, and therefore inform scenario design. For example, a lateral automation feature that requires longitudinal automation to activate may have different use patterns than a lateral feature that can activate alone. HMI design considerations, such as RTI characteristics (e.g., multi-modal), cluster display, and type (e.g., screen size, head-up vs. head down, etc.) are likely to differ across different system designs (see e.g., this Handbook, Chapter 15). Different monitoring methods may allow for a sequence of activities that "defeat" the monitoring system or provide an avenue for improper use of the automation (e.g., steering torque detection vs. driver gaze detection; see also, this Handbook, Chapter 11).

Although a critical step, there is no one right way to go about characterizing the driving system; the approach to characterization will vary based on the overall intent of the tests. Essentially, individuals evaluating a system should let their research questions guide the level and depth of characterization needed. The approach laid out in this chapter should be considered only as a starting point. Guidelines have been put forward for characterizing interface designs for driving automation (Campbell et al., 2018), which were based on the output from early studies of Level 2 and 3 systems (Blanco et al., 2015).

22.1.2.1 Automated Features

First and foremost, in terms of characterization, is the type of driving automation features that are present in a particular platform (also see this Handbook, Chapter 2). As described in SAE J3016, a vehicle may have multiple features that operate at different levels of automation in different combinations of activation or different operational driving domains (ODDs). There are any number of other items that may be of interest to a particular test or research question; a short example characterization is included in Table 22.1. The type of feature, speed ranges of activation, methods of activation, and whether or not the feature can activate alone are all important details for characterization. Although included in the table as Level 3, low-speed traffic jam assist features may or may not be classified as Level 3 by the manufacturer, which could even vary based on location.

For commercially available systems, a great deal of information may be contained in manufacturer sources (e.g., owner's manuals, manufacturer's website); this may include the specified level of automation for a system. As part of characterization, experimenters should operate systems and verify that the published specifications

TABLE 22.1
Example Characterization of Features for Level 1, Level 2, and Limited Level 3 Capability

Feature	Lateral/ Longitudinal	Speed	Activation Method	Can Be Activated Alone	SAE Level	Alerts
Adaptive cruise control	Continuous longitudinal support		Steering wheel button	Yes	1	FCW
Lane centering	Continuous lateral support	Above 40 mph	Automatic when speed is crossed; system setting to disable	No—Requires ACC	2	RTI
Traffic jam pilot	Continuous lateral and longitudinal support	Below 35 mph	Steering wheel button; HMI notifies when available	No	3	RTI

for the system are accurate and if there are affordances from the system, such as extended periods of hands-off driving or feature activation on improper road types, which are not within the intended use but are nonetheless available to the driver. This type of characterization helps to determine the most likely avenues of misuse or other improper use cases that may be observed in naturalistic settings or otherwise require testing.

22.1.2.2 Request to Intervene

RTIs are notifications to the driver that the driving automation requires an intervention. To the best of the researchers' abilities, the source or trigger for the RTI should be characterized (e.g., hands-off wheel detection, driver gaze tracking, and/or external conditions). It may be determined that these alerts may be generated by vehicle software based on combinations of different variables. The timing, modality, and the identified trigger for the RTI should be noted for the testing platform, an example is included in Table 22.2.

22.1.2.3 Other Alerts

In some cases, alerts or notifications and driver responses to these alerts will be the primary focus of study. Alerts for novel or new applications, such as connected vehicle alerts should be characterized insofar as the alert triggering conditions, modality, etc. of the alert are known to the researchers. Alert specifications not only allow for interpretation of the data and scenario development but also provide information that may be explained to participants as part of training.

Alerts may not be of primary interest to a testing scenario. However, the nature and presence of alerts should be noted (e.g., FCW, Blind Spot Warnings, Lane Departure Warnings). Again, the triggering conditions and operational ranges, should be understood by the research team in order to provide information to participants as necessary. Particularly for a naturalistic research study, alerts are likely to be encountered by a participant driver.

22.1.2.4 Connected Vehicle Capabilities

In addition to characterizing automated features, the presence and characterization of any connected vehicle technology should be included as part of the system characterization. Connected vehicle technology enables communication between vehicles (vehicle to vehicle, V2V) and between vehicles and infrastructure

TABLE 22.2
Example RTI Characterizations

	Source (Alert Trigger)	Number of Stages	Stage Duration (seconds)	Total Duration (seconds)	Consequence
Level 2 RTI	Steering wheel torque	2	15	30	Lane centering is disabled for duration of trip
Level 3 RTI	External conditions	3	15	45	Vehicle slows to a stop in lane

(vehicle to infrastructure, V2I) using short-range radio frequencies. This communication can be used by various in-vehicle systems and/or driving automation to provide information to the driver/automated system, such as crash warning system (CWS) alerts. V2I features may provide navigation, variable road signage, or other infrastructure-based information to the driver and/or driving automation system (also see Chapters 2 and 19).

Although existing radar and camera base systems can provide information on FCW situations, V2V messaging may allow for other types of CWS alerts, such as a left turn across path warning or an alert that a vehicle is crossing the intersection. This type of alert is likely novel to most drivers; as such aspects of the HMI should be noted. Among the aspects of the HMI to consider are the modalities (visual, auditory, and tactile) of the interface that are used to notify the driver of an impending crash conflict (also see Chapter 15).

22.1.3 Participant Training

No matter how participants are selected (paid volunteers from the surrounding community, university students seeking course credit), it's likely that at least some of them have heard the term "self-driving car" or "autonomous vehicle" used in news reports, YouTube videos, or other media, but have little practical experience with any driving automation. Participants may have already decided that they do not trust driving automation or have a preconceived notion of how well it will work (see also, this Handbook, Chapters 4, 5). A recent survey found 71% of participants were afraid to ride in a self-driving car (AAA Public Affairs, 2019). Generally speaking, participant training should be designed to provide information for a driver to safely and effectively operate the system(s) of interest, and not have participants rely only on these existing notions to guide their use of the automated systems. For example, Russell et al. (2018) created training for naturalistic driving participants to mimic an ideal dealership experience. It should go without saying that system characterization must inform training; providing inaccurate information may bias the results of testing or otherwise lead to improper use of driving automation (see Section 22.1.4).

The goals of the desired test must also be considered; different situations will have different instructional requirements. The level of detail provided in participant instructions or training should be tuned to the expected use case for the systems being tested (e.g., this Handbook, Chapter 18). For example, if the desired scenario is for an unfamiliar driver experiencing automation that unexpectedly activates (e.g., a rental car), providing little or no training to a participant may be necessary. If the test case is not intended to represent a completely unaware driver, providing no information to a participant will probably lead to situations where the participant is confused about how to operate the system or respond appropriately to system alerts or messages. Participants who receive too much (or overly technical) training may not accurately reflect the knowledge of the "typical" user. However, these results may generalize better to commercial vehicle drivers or other highly experienced populations. Likely training should be somewhere in between in most cases: a general description of the features, what they do, how they activate, and what alerts or messages may be displayed to the driver. A test drive where a participant can ask questions and experience the features for the first time is a second step that should help reinforce instructions.

Training considerations may not be limited to the use and activation of automation systems, it may be necessary to provide training for participants on any other unfamiliar or novel aspect of a test, such as a specific non-driving task. Even if the tasks are not novel, the results could be biased if the application of the task is confusing or misunderstood by the participant. Finally, if the goal of the test is to develop a training system or to compare different training methods, designing tests to compare the methods will be required. For example, a written or practical evaluation to determine what information was understood and/or retained from the training material by the participants.

22.1.4 Training and Improper Use

A key question for the rollout of driving automation will be how drivers use and adopt the systems; automation that is improperly used may lead to crashes. As noted previously, there may be scenarios where a system affords the driver behaviors that are outside the design intent, but otherwise possible (e.g., hands-off wheel, performing a non-driving task). Testing these affordances will require deliberate creation of improper use cases, such as asking a driver to perform a non-driving task during periods of automation use in a controlled setting. For example, Blanco et al. (2015) asked participants to respond to emails using an experimenter-provided tablet during Level 2 operation, and asked participants to watch videos during Level 3 operation. Russell, Atwood, and McLaughlin (in press) provided training to participants that described lateral automation capabilities above or below the actual capability (i.e., described a highly capable lane centering system when in reality the capability was more akin to a lane departure prevention system).

22.2 COMMERCIAL VEHICLE TESTING

The potential economic benefits of automated driving systems for commercial vehicles may lead commercial fleets to become early adopters of driving automation. These potential benefits include flexibility in driver's hours of service, reductions in labor costs, reduced liability for crashes, and fuel economy improvements (Poorsartep & Stephens, 2015). Commercial vehicle drivers may not have a choice when adopting driving automation; they may be assigned a tractor trailer that is already equipped with ADAS or other driver assistance technology. Existing research on commercial applications of ADAS systems (e.g., ACC, FCW, and Lane Departure Warning Systems) suggests that the systems still produce some false activations and extraneous low-level feedback which may frustrate drivers or reduce their trust in the systems (Grove, Atwood, Blanco, Krum, & Hanowski, 2017). It was also found that these drivers were willing to use low-level automation such as ACC in adverse weather conditions against manufacturer recommendations (Grove, et al., 2015) and that there was a small, measurable difference in their visual behavior while using ACC and following a lead vehicle (eyes-off-road time was slightly higher; Grove, Soccolich, Engstrom, & Hanowski, 2019). Recently the commercial vehicle industry has begun to recognize issues with driver confusion or frustration towards FCW and lane departure alerts due to variations between brands of systems, generations of system, and integration approaches (Technology & Maintenance Council, 2018). The best practices from these efforts may inform designs for higher levels of automation for commercial vehicle ADAS or commercial AVS (Technology & Maintenance Council, 2019).

In addition to issues with driver acceptance, commercial vehicles also introduce additional challenges for testing and development. Testing of commercial vehicles presents specific challenges in terms of ODDs as well as specialized automation that may be different than driving automation for passenger vehicles. For example, trucking operations that swap trailers may be limited to placing sensors only on the truck/tractor itself, creating visibility challenges unique to commercial vehicles. Commercial vehicles are also subject to roadside inspections, and automated systems may need to interact with state personnel who perform these inspections or their infrastructure. Commercial vehicles may also operate on private properties or non-public roadways as part of their operation where road markings are limited. Considering these issues during testing will be critical to uncovering the edge cases for driving automation for heavy vehicle applications. Additional examples of these challenges specific to heavy commercial trucking are described in the following sections.

22.2.1 Drayage

One example of a non-public roadway usage for commercial vehicles is drayage. Drayage trucks typically work to transfer cargo between mixed modes of transportation, such as offloading cargo from a ship to then be transported by heavy truck or train. These operations are often conducted in a space with controlled access, such as a port of entry or rail yard facility. These characteristics make the domain attractive for heavy vehicle automation; the operation space is confined, there is little to no mixed traffic, there are relatively fixed transit routes, and there are lower speeds of operation (see Smith, Harder, Huynh, Hutson, & Harrison, 2012, for an overview of drayage operations and facilities). Drayage may also offer unique challenges to automated commercial vehicles as roadway infrastructure could differ significantly from public roads. Additionally, while most driving may take place in areas closed to public traffic, there may be some driving on public roads in order to get the cargo to a nearby destination for further transit. These transitions between private and public space may provide test cases of transfer of control between automated driving and manual driving.

22.2.2 Platooning

Another application of automation that may be unique to commercial vehicles is platooning. Platooning involves a series of vehicles following each other at relatively short headways for extended periods of time. Platooning is typically accomplished by V2V communication in order to synchronize throttle or steering and assist a driver, but future implementations could also include higher levels of automation which could operate without a driver present. Platooning offers the opportunity for the following vehicles to reduce airflow drag and improve fuel economy at higher speeds, which presents an economic opportunity in commercial vehicle operations that drive long distances at highway speeds. However, the following distances necessary to achieve optimal gains are expected to be relatively short, as low as 10 feet (Tsugawa, Jeschke, & Shladover, 2016) and as high as 50 feet (Lammert, Duran, Diez, & Burton, 2014).

The short headways necessary for platooning reduce the time in which the following vehicles can react and create blind spots in front of the following vehicles in which sensors cannot see beyond the leading vehicle. This can be overcome with communication between leading vehicles and following vehicles, but the lead vehicle may need to consider whether there is a platoon following it in choosing how to react to potential conflicts. Additionally, the points at which a human driver or the automation needs to engage/disengage a platoon are another area of potential testing unique to commercial vehicles. Depending on how platoons are designed to form or dissolve, they may involve a human that is in control at what would typically be considered an unsafe headway or the automation must transition over time from an independent state at a longer headway to a synchronized state at a shorter headway (or vice versa). These handover or transition points could lead to edge cases and testing requirements that are unique to commercial vehicles.

A final concern for platooning testing is how the systems should react to vehicles around a platoon. Light vehicles may attempt to cut between platooning trucks (depending on the following headway required for a platoon), and platoons may impact how surrounding vehicles attempt to enter, exit, or change lanes on a highway. Anticipating these behaviors around platoons and including them in testing scenarios will be critical for ensuring safe deployment in the future.

22.3 TESTING EARLY IN DEVELOPMENT: DATA ANALYSIS AND DRIVING SIMULATION

For the purposes of this chapter, early stages of testing are those prior to development of a working prototype in a moving vehicle. Data analysis and modeling along with simulator testing are reviewed briefly, as they relate to participant testing. These two methods can be implemented in an iterative fashion to improve feature performance before a prototype system is built. For example, consider a traffic scenario that may be familiar to the reader; a slowed vehicle is revealed to a driver. As outlined in Figure 22.2, three vehicles are traveling down a roadway (Panel 1), when the lead vehicle brakes. In certain circumstances, this slow/stopped vehicle cannot be seen by the driver of the third vehicle. When the middle vehicle changes lanes to avoid the slowed lead vehicle (Panel 2), the slow/stopped vehicle is revealed to the third driver who must then respond, potentially with an evasive maneuver (e.g., hard braking; panels 3 & 4).

As part of the development of a longitudinal automation feature, examples of this type of interaction can be selected from existing naturalistic driving datasets (e.g., the Second Strategic Highway Research Program (SHRP2); Dingus et al., 2014). Using the video, acceleration, global positioning system (GPS), and other data collected as part of the study, the various actors within the scenario can be modeled in a computer-based simulation (e.g., CarSim, MATLAB, or other software program).

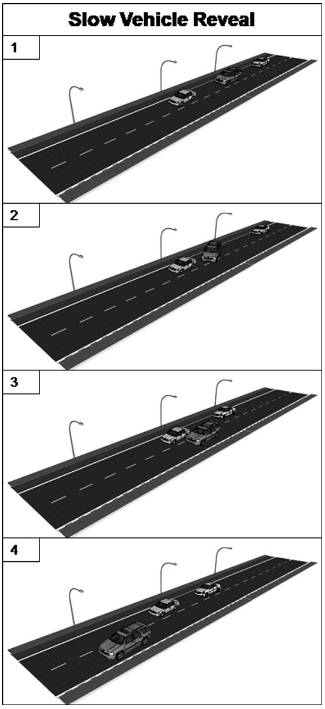

FIGURE 22.2 Slowed vehicle reveal scenario.

Automated features or other ADAS systems can then be modeled (for example, implementing a connected vehicle application that detects the lead vehicle slowing in the model). These modeled features can then be implemented in a driving simulator to test the features with a naïve driver; the driver response data can then be used to further refine model parameters, leading to additional tests.

22.3.1 Naturalistic Driving Data for Scenario Development

Testing of automated driving systems will require millions if not billions of miles to determine overall reliability and safety which would take tens or hundreds of years to achieve based on a recent RAND report (Kalra & Paddock, 2016). Analysis of the millions of miles of existing naturalistic driving data may speed up the testing process. The following chapter will cover the analysis of large datasets in further detail (this Handbook, Chapter 23); however, data analysis and modeling techniques are included here for their contributions to early stage testing and development of automated features. Mining existing datasets, determining points of comparison, and developing models and simulations of features can provide a valuable and easily repeatable method for testing features in early stages of development.

Although the end goal of driving automation systems may be to replace the human driver (or at least to augment the driver), automated systems will first be compared to the human driver they intend to replace. A reasonable place to start is to use existing data on typical drivers, namely naturalistic driving data, such as that found in the SHRP 2 dataset (Dingus et al., 2014). Most approaches to naturalistic data have been to study rates of safety-critical events (SCEs; crashes and near crashes) along with the factors present during each event.[1] The remaining data are used to select baseline cases without crashes for computing odds ratios and other analyses. Crashes are rare events overall, and the majority of the data are uneventful in the conventional sense. However, this "leftover" data provide a multitude of driving scenarios that could be informative for automated system development and testing. Essentially, system function can be compared to data from actual drivers in the scenario; as noted previously these scenarios can then be modeled in a computer-based simulation and system parameters tuned to desired performance specifications prior to testing with a human driver. This is an emerging approach for driving automation system development, and most work is being conducted in the private realm but is conceptually similar to hardware in the loop simulation (see e.g., Murray-Smith, 2012).

To what standard should a system be tested? Should a system be developed to be similar to the mean following distance? What about following closer? Why not choose one of the outliers for an added margin of safety? There are no right answers to any of these questions; but as a researcher, decisions will need to be made about how the system should be tested, and this decision will have an impact on the outcome of the test. There are a wide variety of driving styles; and driving style is going to change depending on the surrounding traffic, weather and environmental factors, and so on. This makes even what is on the face a simple question, such as how closely should a longitudinal feature follow lead traffic, more challenging. At some point in development, it has to be determined what type of driving the automated system is going to achieve. Figure 22.3 shows boxplots calculated from following distance

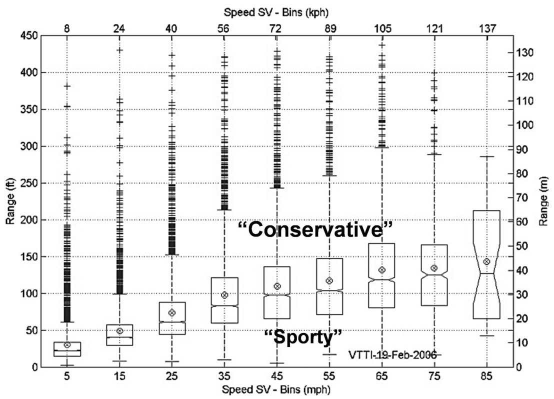

FIGURE 22.3 Boxplots of mean following distance data for 10,000 trips from the 100-car NDS with range (*Y* axis) and speed (*X* axis) in metric and standard units.

[1] Note that some controversy surrounds the use of near crashes to understand the factors that influence crashes (Guo, Klauer, McGill, & Dingus, 2010; Knipling, 2015)

observed during the 100-car NDS (Dingus et al., 2006). Data are shown for 10,000 trips, binned across the speed ranges on the *x*-axis. Based on the following distance classification, those on the higher ranges could be classified as "conservative" with the shorter following distances considered "sporty."

Consider the example of a slowed vehicle reveal scenario (Figure 22.2) and designing vehicle automation that can resolve the scenario safely. Using this type of driving data, iterative models of different following distance envelopes can be used to narrow down design considerations (e.g., following at a safe distance to resolve a vehicle reveal). Again, analysis methods for large datasets are covered in further detail in other chapters (this Handbook, Chapter 23). For the purpose of iterative testing, upon narrowing the feature parameters, testing with human operators in a driving simulator is likely appropriate.

22.3.2 Driving Simulator Testing

The majority of published tests using human subjects and vehicle automation utilize a driving simulator of some type. Driving simulator options run the gamut from simple monitors on a tabletop computer, large screens with a typical vehicle interior setup, to immersive driving simulators with actual movement (instead of visual-only motion cues) (see Weir, 2010, for an overview driving simulator applications in relation to HMI development). Testing using a driving simulator has some advantages, namely cost and logistics, when compared with testing on live roads. Re-programming of simulations allows a researcher to collect data for a wide variety of different conditions, features, scenarios, and so on.

Aside from issues of motion sickness (see Brooks et al., 2010, for an overview of simulator sickness and potential detection and mitigation), simulators pose little safety threat to a participant. However, this can be both a blessing and a curse for testing. On one hand, a crash imminent scenario can be implemented without any real danger; but without any consequence for an incorrect decision a driver's response may not be representative of what might be observed on real roads (Ranney, 2011). The key then is making sure the test of interest is matched with the simulator. Specifically, the fidelity of the simulator and the accuracy at which automated features are replicated need to be matched. Even if there is some non-visual motion cue, it is likely that it does not have the same acceleration profile as a vehicle traveling on an actual road surface. This lack of a realistic lateral motion cue may delay driver responses to lateral automation problems such as lane drifts (Kaptein, Theeuwes, & Van Der Horst, 1996).

Since the test environment is programmed at the direction of the experimenter, it certainly allows for more overall experimenter control than naturalistic data collection. The scenarios can play out exactly the same way for different participants. This may be particularly useful in early stages of system development or testing, in which individual aspects of a system are being tested iteratively and rapidly (e.g., alert timings, tones, icons). Physical configurations and layout changes can also be implemented iteratively in a simulator where mistakes made have little consequence. Still, at some point, testing must move onto live road testing using vehicle-based driving automation. These early steps should establish design guidance for on-road system development.

22.4 MID-LATE TESTING: ON-ROAD EXPERIMENTS

Early stages of testing are essential for development and refinement of system parameters, physical layout of controls and displays, HMI messages and timing, and so on. Early and iterative testing approaches should lead to a better initial design of a testable prototype system. Even based on the optimum modeling and simulator tests, a system will require further testing on live roadways (including both test track and public roads). There may be additional mechanical system limitations that arise only when the automated systems are implemented in a prototype vehicle, such as steering system torque limitations, machine vision limitations when reading lane markings, and so on.

As noted above, testing at this stage will involve on-road testing using driving automation. The focus of this section is on scenario selection and on one particular method of on-road testing, the Wizard of Oz (WO) approach. Similar to analysis of existing naturalistic driving data and simulator testing, scenario selection is critical for on-road testing. Although on-road testing can be conducted with commercially available vehicles (Endsley, 2017; Banks, Eriksson, O'Donoghue, & Stanton, 2018), the WO approach is included as it will be a critical path for testing automated driving technologies that are not commercially developed (e.g., Levels 3 and 4 driving automation).

22.4.1 Scenario Selection

Some on-road scenarios may have already been selected from previous data analysis and/or extensions of driving simulator tests (e.g., the slowed vehicle reveal example, Figure 22.2). At this stage of testing, cost and logistics will limit a researcher's ability to iteratively revise and re-test many scenarios. As such, testing on live roadways (either test track or public roads) requires the researcher to narrow down the types of scenarios to those of most interest.

Tests of driver interactions with prototype or WO automated systems may be conducted on public roads, depending on the safety and the legality at the state and local level. Essentially, a participant operates an automated system on public roads while one or more researchers provide instruction, analyze behaviors, record participant preferences, and so on. Manipulations may include changes to system characteristics, such as HMI displays and/or messaging strategies (see Blanco et al., 2015), and/or automated system capabilities (see Russell et al., in press). This approach will provide data for driver interaction with automated systems in typical settings, with traffic conditions that will vary over the course of testing. This approach is akin to a quasi-naturalistic study; an experimenter is present, but traffic scenarios develop as they would occur normally in the testing environment.

It may not be an effective use of time or resources to "wait and see" what scenarios develop; a researcher may need to create or orchestrate a specific scenario. Multiple traffic scenarios, environmental variables, and other sources of scenario complexity were noted in the introductory section to this chapter, per SAE J3018. Additional aspects of the automated systems (e.g., lateral features, alert severity, etc.) add another layer of complexity. Given the sheer number of possibilities, it will certainly be helpful to narrow down options. One potential scenario classification method is based on how quickly a response is needed (i.e., urgency), whether the scenario is expected or not (i.e., predictability), and the consequences of not responding or intervening (i.e., criticality) (Gold, Naujoks, Radlmayr, Bellem, & Jarosch, 2017).

Again, there is no single test or even set of tests that should always be conducted; testing will vary based on the feature of interest. However, the highest return on investment may be a "worst-case scenario" approach. Using the classification above, this would be a scenario with high urgency, low predictability, high criticality, assuming it can be implemented safely. These scenarios are dependent on the capabilities of the system or feature of interest, but some examples may include an unresponsive driver (asleep, impaired), major system malfunction (improper steering, sensor failure), or a driver performing a non-driving task during an RTI. This method is dependent on the researcher's ability to perform these tests safely but can yield valuable data on driver responses to these critical situations.

22.4.2 WO Approaches

Although likely familiar to most readers, WO testing refers to testing a system that operates as though it were automated but is not. WO "automation" is achieved through mechanical intervention (e.g., secondary controls), using only pre-determined routes, GPS guidance, or similar methods. This approach may be necessary for a variety of reasons; it is possible that the researchers are interested in testing a general category of automation (e.g., Level 2) rather than a specific platform or system and would therefore avoid using a stock testing platform. Alternately, the state of development or other testing limitations may be such that the system of interest is not ready for deployment but may work on a test track with a pre-programmed path. Alternatively, a lateral control feature may be classified as Level 2, and is capable of maintaining lane position but requires the driver to maintain steering input; adding a mechanical steering system from a confederate driver in a rear seat may allow for testing of Level 3 scenarios on live roads.

As a final note, the WO testing is a flexible testing approach not limited to driver testing. For example, the approach can be used to test interactions between Level 4 driving automation and other road users. Rather than augmenting vehicle technology, a driver may be hidden from view of anyone outside the vehicle. The "ghost driver" (Rothenbucher, Li, Sirkin, Mok, & Ju, 2015) method can give the appearance a vehicle is autonomous, allowing testing for general reactions from other road users and pedestrians, such as responses to signal lights (Ford Motor Company, 2017). Responses can be gathered via post-exposure interviews or video-based reduction of observer response.

22.5 LATE STAGE TESTING: NDSS

For a moment, assume the best-case scenario: A feature was designed using existing naturalistic driving data; models of the feature were tested in computer-based simulation, initial concepts of this feature were then refined iteratively in a driving simulator, and finally a functioning prototype of the feature was subsequently tested in a series of on-road experiments. As discussed, many interesting scenarios were likely not tested for logistical and budgetary reasons. There may be unintended consequences or other unforeseen edge cases that may only be observed when deployed in real-world conditions. While not guaranteed to reveal new edge cases, testing new technology via an NDS may reveal new scenarios of interest as well as provide a test bed for existing assumptions about feature use and adoption. Finally, NDSs may provide insight into changes in driving automation use over time; a long experiment may last several hours, and use of a system for days, months, or even years is likely necessary to test how automation use patterns change over time (see also this Handbook, Chapter 12).

Methods and results for two different NDSs using commercially available Level 2 systems have been recently published (Russell et al., 2018; Fridman et al., 2019). Russell et al. (2018) deployed five different makes and models equipped with Level 2 features to 120 participants, two of each model were listed below:

- 2017 Audi Q7 Premium Plus 3.0 TFSI Quattro with Driver Assistance Package
- 2015 Infiniti Q50 3.7 AWD Premium with Technology, Navigation, and Deluxe Touring Package
- 2016 Mercedes-Benz E350 Sedan with Premium Package, Driver Assistance Package
- 2015 Tesla Model S P90D AWD with Autopilot Convenience (software version 8.0)
- 2016 Volvo XC90 T6 AWD R with Design and Convenience Packages

Vehicles were instrumented with a data acquisition system including camera views of the driver and forward roadway, and accelerometer and GPS data. Each participant drove the vehicles for four weeks. Participants received an introduction to the driving automation, including a test drive before their participation period. A total of 216,585 miles were driven, with 70,384 miles driven with both lateral and longitudinal features active. Fridman et al. (2019) recruited drivers who were already owners of Level 2 capable systems (specifically Tesla drivers), with analyses reported for 21 vehicles and 323,384 total miles (112,427 miles driven with Level 2 active). Cameras were used to record driver behavior and the forward roadway.

The intent here is not to provide a "how to" for the logistics and data collection procedures for conducting an NDS, but some primary issues will be reviewed briefly. The focus in this section is on the issues relating to the participant and data collection, data sampling, and data reduction for testing of automated driving systems. Researchers will need to plan for what types of systems and participants will be studied (e.g., owners of candidate systems or participants to which systems will be loaned). Researchers will also need to decide on what instrumentation (e.g., aspects of the data acquisition system: cameras, accelerometers, GPS, vehicle network information, storage capacity) should be deployed. Likely dedicated staff will be needed to monitor the data collection, to replace and/or re-align cameras, and to manage data as data storage fills up during the data collection period. For a more detailed review of methods for field operational tests and NDSs, see the overview published by the FESTA Consortium (2017).

22.5.1 Participant Selection

Participant selection for NDSs has a few distinct differences from a test track study. The SHRP 2, the largest NDS as of 2019 (32 million miles, ~3,500 drivers), recruited volunteers willing to have their own personal vehicles instrumented with cameras and other data collection equipment (Dingus et al., 2014). This approach to recruiting is dependent on volunteers that already own a vehicle with automated driving capabilities and are then willing to have it instrumented with cameras and other data collection equipment. As such, using this method for testing may be limited due to lack of market penetration and/or finding willing volunteers.

As an alternative, vehicles that include automated driving systems may need to be procured and loaned to participants for some length of time, as was done for the Level 2 NDS (L2NDS; Russell et al., 2018). For either approach, limited budget and funds will call for tradeoffs in the study length and the number of participants selected. Depending on the specific needs of the test, it may be better to conduct a study with longer exposure per driver with fewer participants, or it may be better to collect data for a larger number of participants for a shorter time. For example, longer exposures may be more likely to reveal driver adaptations and habituation to using driving automation (this Handbook, Chapter 12). If the approach is to loan vehicles to participants, then training methods for participants (as described in the introduction section) should be considered before their data collection period begins.

22.5.2 Data Sampling & Reduction

The hallmark of an NDS is that data are recorded continuously any time that the vehicle is in operation. This amounts to an enormous volume of data for even a single drive taken by a single participant, let alone across an entire dataset that consists of trips taken over months or years. The typical approach is to sample the continuously recorded data for subsequent reduction and analysis. The number of samples, the duration of each sample, and types of samples (i.e., the sampling strategy) will vary based on the size of the dataset and the goals of the analysis. The events to be sampled will be tailored to the needs of the particular test but will likely include specific alerts or other transitions points of the driving automation (e.g., RTI alerts issued by the automated features, or automation activation or deactivation). It is noted that alternative methods that use machine vision and other neural net analytics (e.g., Fridman et al., 2017) to analyze data continuously (as opposed to discrete samples) are under development.

22.5.2.1 Data Sampling

Sampling strategies may require comparisons between levels of automation, particularly for driving automation that can operate at different levels (depending on which features are activated). Baseline selection may be somewhat tricky, and random sampling of "non-eventful" driving may lead to results that are not representative of the full range of capabilities a system may have. Aside from excluding alerts or other situations of interest, a sampling plan should be based on information from the characterization process, in order to select samples that include all levels of

TABLE 22.3
SCE Descriptions as Used in SHRP 2

Severity Level	Description
Most severe	Any crash that results in any injury requiring medical attention, or one that includes an airbag deployment or requires vehicle towing
Police reportable crash	A crash that does not meet the requirements for a Level I crash, but does include sufficient property damage that warrants being reportable to the police
Minor crash	A crash that does not meet the requirements for a Level II crash, but does result in minimal damage
Low risk crash	Tire/curb strike
Near crash	Any circumstance that requires a rapid evasive maneuver by the subject vehicle, or any other vehicle, pedestrian, cyclist, or animal, to avoid a crash

automation for which a system is capable and appropriate ODDs for different levels of automation. For comparisons between automated driving and non-automated (or non-assisted) driving, samples can be taken from the same ODDs with and without features activated. If the automated driving system has one or more of those features that can be activated separately, samples should be taken from each "level" of activation. For example, naturalistic studies of vehicles equipped with both lateral and longitudinal automated features sampled Level 0 (no features active), Level 1 (one feature active), and Level 2 (both features active) during similar driving scenarios (e.g., highway driving above 40 mph; Russell et al., 2018).

SCEs consisting of crashes and near crashes will certainly be of interest in most if not all test cases. If not reported to the research team directly (by participant drivers themselves or by law enforcement) SCEs can be detected by kinematic data and subsequently confirmed by reviewing video data (see data reduction section). Table 22.3 shows the definitions of SCEs used in the SHRP 2 study (Antin et al., in press).

As a final note, driver behaviors themselves do not automatically elevate a sample to an SCE, even if the behavior is egregious (e.g., visibly intoxicated, sleeping). This also applies to the state of automation in relation to driver behavior (e.g., texting while driving with the automated feature).

22.5.2.2 Data Reduction

Data reduction, sometimes referred to as data annotation, is a step prior to analysis in which each sample epoch is reviewed and the relevant driver, vehicle, and environmental factors are classified by a researcher following a specific protocol. Typically, the time-synchronized vehicle data (e.g., automation state, driver inputs, etc.), sensor data (accelerometer, GPS, etc.), and video data are reviewed frame by frame as necessary (based on the recording rate of the video). Relevant driver factors include presence of non-driving tasks, driver gaze patterns, hands-on-wheel behaviors, pedal behaviors (e.g., accelerator release, brake presses, etc.), and any visible signs of impairment, etc. Driver behaviors that occur with and without driving automation, including the presence of non-driving tasks will likely be of importance to any naturalistic study.

Describing the driving scenario is an intricate process. In addition to the driver factors, environmental factors should also be annotated. For example, the researcher should describe the scenario in relation to the bulleted list of driving scenarios and domain descriptions listed in the introduction of this chapter (a modified list follows).

- Roadway Type
  - Limited access freeway
  - Highway (single or multi-lane),
  - Arterial roads
  - Residential streets
  - Driveway, parking lot, or structure
- Traffic Environment
  - Level of service
  - Vehicles present (e.g., leading and adjacent vehicles)
  - Pedestrians
  - Animals
  - Relation to intersection
  - Traffic control devices (signals, signs, curbs, guardrails, etc.)
- Time of Day & Lighting Conditions
  - Day
  - Night (unlit)
  - Night (lit)
  - Dusk
  - Dawn
- Inclement weather present
  - Rain
  - Snow
  - Fog

Specific operation definitions used for the analysis should be compiled into a "data dictionary" for each study. The data dictionary includes operational definitions used to reduce each scenario that can be referenced by any reductionist at any time and can be used to train new reductionists. An example data dictionary can be consulted to guide protocol development and descriptions of the vehicle, driver, and environmental factors present in the scenario (Virginia Tech Transportation Institute, 2015). Other sources for operational definitions and calculating driver performance metrics include Green (2012) and SAE International (2015b).

22.6 CONCLUSION

We sit at a critical transition point for driving automation deployment and testing. Testing of Level 4 and higher automated driving systems is currently underway by multiple companies (e.g., Waymo, Cruise); however, the many engineering challenges for autonomous driving leave timelines for deployment are still uncertain.

Characterization of these Level 4 systems, including their expected ODDs as well HMIs, will be critical for testing these systems. ADAS (e.g., Level 1) and partial driving automation (e.g., Level 2) are becoming available in more affordable makes and models. Early test track studies of these systems conducted with prototype vehicles (Blanco et al., 2015) have provided understandings and underpinnings of human factors guidance for driving automation development (Campbell et al., 2018; see also, this Handbook, Chapter 15). Additional research is likely underway in the private realm, as OEM and Tier 1 suppliers test and develop systems as part of system and product development.

As driving technology continues to improve, a question remains as to the degree to which drivers will be monitoring systems as opposed to actively driving with automation. As noted above, two naturalistic studies of Level 2 driving automation use have recently been published. The results for these studies indicate that drivers using Level 2 automation were using driving automation in a largely active fashion; there was no difference in non-driving task prevalence observed by Russell et al. (2018) nor were more dangerous tasks more likely while Level 2 was active (e.g., texting or browsing on a smartphone) nor any relationship between Level 2 use and SCEs. Fridman et al. (2019) specifically looked at periods of de-activation of Level 2 states, finding 90.6% of deactivations categorized as anticipatory in nature. It will be of great interest to see if similar results are observed with Level 3 and higher systems.

As automated driving technology continues to become more prevalent, testing and development will be critical to the safe deployment of the technology. Again, the goal for testing should be of practical significance; when evaluating driving automation technology, the end result may be one when "nothing happens." Reaction times or distractions observed in early stages should be tested in naturalistic settings to determine if there is an overall safety risk associated. As on-road and naturalistic testing with vehicle automation continue, we should not be surprised as results emerge that seem counterintuitive when compared with findings from simulator or on-road experimental findings.

Finally, the testing approaches in this chapter were largely focused on system development and evaluation with a targeted focus on driver testing. As automated driving of all levels becomes more widespread, testing must continue to interrogate broader effects of automated driving at the "transportation system" level (this Handbook, Chapters 13, 19). Interactions with other "non-automated" driving in a mixed fleet will likely lead to novel cases that cannot be easily predicted until they emerge. Similarly, interactions with pedestrians, cyclists, and other vulnerable road users at a broader scale should be optimized as informed by testing and evaluation.

ACKNOWLEDGMENTS

This chapter was heavily informed by the authors conducting funded studies from the National Highway Traffic and Safety Administration as well as other proprietary sponsors. A special thanks to the editors and other reviewers for valuable feedback on early drafts of this chapter.

REFERENCES

AAA Public Affairs. (2019). *Three in Four Americans Remain Afraid of Fully Self-Driving Vehicles*. Retrieved May 12, 2019, from https://newsroom.aaa.com/tag/autonomous-vehicles/

Antin, J. F., Lee, S., Perez, M. A., Dingus, T. A., Hankey, J. M., & Brach, A. (in press). Second strategic highway research program naturalistic driving study methods. *Safety Science, 119,* 2-10.

Banks, V. A., Eriksson, A., O'Donoghue, J., & Stanton, N. A. (2018). Is partially automated driving a bad idea? Observations from an on-road study. *Applied Ergonomics, 68*, 138–145.

Blanco, M., Atwood, J., Vazquez, H. M., Trimble, T. E., Fitchett, V. L., Radlbeck, J. … Morgan, J. F. (2015). *Human Factors Evaluation of Level 2 and Level 3 Automated Driving Concepts* (DOT HS 812 182). Washington, DC: National Highway Traffic Safety Administration.

Brooks, J. O., Goodenough, R. R., Crisler, M. C., Klein, N. D., Alley, R. L., Koon, B. L., … Wills, R. F. (2010). Simulator sickness during driving simulation studies. *Accident Analysis & Prevention, 42*(3), 788–796.

Campbell, J. L., Brown, J. L., Graving, J. S., Richard, C. M., Lichty, M. G., Bacon, L. P., … Sanquist, T. (2018). *Human Factors Design Guidance for Level 2 and Level 3 Automated Driving Concepts* (DOT HS 812 555). Washington, DC: National Highway Traffic Safety Administration.

Dingus, T. A., Hankey, J. M., Antin, J. F., Lee, S. E., Eichelberger, L., Stulce, K., … Stowe, L. (2014). *Naturalistic Driving Study: Technical Coordination and Quality Control* (SHRP 2 Rep. S2-S06-RW-1). Washington, DC: National Academies.

Dingus, T. A., Klauer, S. G., Neale, V. L., Petersen, A., Lee, S. E., Sudweeks, J., & Knipling, R. R. (2006). *The 100-Car Naturalistic Driving Study, Phase II Results of the 100-Car Field Experiment* (DOT HS 810 593). Washington, DC: National Highway Traffic Safety Administration.

Endsley, M. (2017). Autonomous driving systems: A preliminary naturalistic study of the Tesla Model S. *Journal of Cognitive Engineering and Decision Making, 11*, 225–238.

FESTA Consortium. (2017). *FESTA Handbook Version 7.* Retrieved from https://fot-net.eu/Documents/festa-handbook-version–7/

Ford Motor Company. (2017). *Ford, Virginia Tech Go Undercover to Develop Signals That Enable Autonomous Vehicles to Communicate with People*. Retrieved from https://media.ford.com/content/fordmedia/fna/us/en/news/2017/09/13/ford-virginia-tech-autonomous-vehicle-human-testing.html

Fridman, L., Brown, D. E., Glazer, M., Angell, W., Dodd, S., Jenik, B., … Abraham, H. (2017). MIT autonomous vehicle technology study: Large-scale deep learning based analysis of driver behavior and interaction with automation. *arXiv:1711.06976*. doi: 10.1109/ACCESS.2019.2926040

Fridman, L., Brown, D., Kindelsberger, J., Angell, L., Mehler, B., & Reimer, B. (2019). *Human Side of Tesla Autopilot: Exploration of Functional Vigilance in Real-World Human-Machine Collaboration*. Cambridge, MA: MIT. Retrieved from https://hcai.mit.edu/tesla-autopilot-human-side.pdf

Gold, C., Naujoks, F., Radlmayr, J., Bellem, H., & Jarosch, O. (2017). Testing scenarios for human factors research in level 3 automated vehicles. *International Conference on Applied Human Factors and Ergonomics* (pp. 551–559). Berlin, Springer.

Green, P. (2012). Standard definitions for driving measures and statistics: Overview and status of recommended practice J2944. *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 28–30). Eindhoven, The Netherlands.

Grove, K., Atwood, J., Blanco, M., Krum, A., & Hanowski, R. (2017). Field study of heavy vehicle crash avoidance system performance. *SAE International Journal of Transportation Safety, 5*(1), 1–12.

Grove, K., Atwood, J., Hill, P., Fitch, G., DiFonzo, A., Marchese, M., & Blanco, M. (2015). Commercial motor vehicle driver performance with adaptive cruise control in adverse weather. *Procedia Manufacturing, 3*, 2777–2783.

Grove, K., Soccolich, S., Engstrom, J., & Hanowski, R. (2019). Driver visual behavior while using adaptive cruise control on commercial vehicles. *Transportation Research Part F: Traffic Psychology and Behavior, 60*, 343–352.

Guo, F., Klauer, S., McGill, M., & Dingus, T. (2010). *Evaluating the Relationship Between Near-Crashes and Crashes: Can Near Crashes Serve as a Surrogate Safety Metric for Crashes?* (DOT HS 811 382). Washington, DC: National Highway Traffic Safety Administration.

Kalra, N. & Paddock, S. M. (2016). Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability? *Transportation Research Part A: Policy and Practice*, *94*, 182–193.

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

Knipling, R. (2015). Naturalistic driving events: No harm, no foul, no validity. *Proceedings of the Eighth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*. Iowa City: University of Iowa.

Lammert, M., Duran, A., Diez, J., & Burton, K. (2014). Effect of platooning on fuel consumption of Class 8 vehicles over a range of speeds, following distances, and mass. *SAE International Journal of Commercial Vehicles, 7*(2), 626–639.

Lee, J. D., Wickens, C. D., Liu, Y., & Boyle, L. N. (2017). *Designing for People: An Introduction to Human Factors Engineering*. Charleston, SC: CreateSpace.

Murray-Smith, D. (2012). *Modelling and Simulation of Integrated Systems in Engineering*. Philadelphia, PA: Woodhead Publishing.

Poorsartep, M. & Stephens, T. (2015). Truck automation opportunities. In G. Meyer, & S. Beiker (Eds.), *Road Vehicle Automation 2. Lecture Notes in Mobility*. Berlin: Springer.

Ranney, T. (2011). Psychological fidelity: Perception of risk. In D. Fisher, M. Rizzo, J. Caird, & J. Lee (Eds.), *Handbook of Driving Simulation for Engineering, Medicine and Psychology*. Boca Raton, FL: CRC Press.

Rothenbucher, D., Li, J., Sirkin, D., Mok, B., & Ju, W. (2015). Ghost driver: A platform for investigating interactions between pedestrians and driverless vehicles. *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. ACM, New York, NY, USA, pp. 44–49.

Russell, S. M., Atwood, J., & McLaughlin, S. (in press). *Driver Expectations for System Control Errors, Engagement, and Crash Avoidance During Level 2 Driving Automation*. Washington, DC: National Highway Traffic Safety Administration.

Russell, S. M., Blanco, M., Atwood, J., Schaudt, W. A., Fitchett, V. L., & Tidwell, S. (2018). *Naturalistic Study of Level 2 Driving Automation Functions* (DOT HS 812 642). Washington, DC: National Highway Traffic Safety Administration.

SAE International. (2015a). *Guidelines for Safe On-Road Testing of SAE Level 3, 4, and 5 Prototype Automated Driving Systems*. Warrendale, PA: Society for Automotive Engineers.

SAE International. (2015b). *Operational Definitions of Driving Performance Measures and Statistics*. Retrieved from https://saemobilus.sae.org/content/J2944_201506/

SAE International. (2016). *Surface Vehicle Recommended Pratice J3016: Taxonomy and Definitions for Terms Related to driving Automation Systems for On-Road Motor Vehicles*. Warrendale, PA: Society for Automotive Engineers.

Smith, D., Harder, F., Huynh, N., Hutson, N., & Harrison, R. (2012). Analysis of current and emerging drayage practices. *Transportation Research Record, 2273*, 69–78.

Technology & Maintenance Council. (2018). Technological advances in next generation collision warning driver interfaces. *The Trailblazer: The Technical Journal of TMC's 2018 Annual Meeting.* ATA, Arlington, VA, USA,. 40–45.

Technology & Maintenance Council. (2019). RP 430 update. *The Trailblazer: The Technical Journal of TMC's 2019 Annual Meeting.* ATA, Arlington, VA, USA, 39–40.

Tsugawa, S., Jeschke, S., & Shladover, S. E. (2016). A review of truck platooning projects for energy savings. *IEE Transactions on Intelligent Vehicles, 1*(1), 68–77.

Virginia Tech Transportation Institute. (2015). *SHRP 2 Researcher Dictionary for Video Reduction Data Version 4.1.* Blacksburg, VA: VTTI.

Weir, D. H. (2010). Application of a driving simulator to the development of in-vehicle human-machine-interfaces. *IATSS Research, 34*, 16–21.