FOUNDATION FOR TRAFFIC SAFETY

SAFETY RESEARCH USING SIMULATION
UNIVERSITY TRANSPORTATION CENTER

EMERGING TECHNOLOGIES
TECHNICAL REPORT

Identifying Outcome Measures to Evaluate Drivers' Knowledge of Vehicle Automation

OCT 2024

607 14th Street, NW, Suite 201
Washington, DC 20005
202-638-5944
AAAFoundation.org

Title

Identifying Outcome Measures to Evaluate Drivers' Knowledge of Vehicle Automation

(October 2024)

Authors

Apoorva Hungund[1], Jessica McDonough[2], Meng Wang[1], Austin Svancara[2], Shannon C. Robert[1], Anuj K. Pradhan[1], & William J. Horrey[2]

[1] *University of Massachusetts–Amherst*

[2] *AAA Foundation for Traffic Safety*

Author ORCID Information

Hungund:	https://orcid.org/0000-0001-7116-0165
McDonough:	https://orcid.org/0009-0004-1836-1995
Wang:	https://orcid.org/0000-0002-3304-0610
Svancara:	https://orcid.org/0000-0002-7137-219X
Roberts:	https://orcid.org/0000-0002-0052-7801
Pradhan:	https://orcid.org/0000-0002-7612-4208
Horrey:	https://orcid.org/0000-0002-9533-4411

Foreword

One of research focus areas for the AAA Foundation for Traffic Safety is 'Emerging Technologies,' which examines drivers' understanding, acceptance, and use of emerging vehicle technologies, the benefits of their use, as well as risks that might arise from misunderstanding or misuse of these technologies. The availability of new technologies continues to expand in production vehicles and in the U.S. fleet. Features such as adaptive cruise control and lane centering assist are taking on more active control of driving tasks under drivers' supervision. However, past research has shown that drivers often misunderstand the purpose, capabilities, and limitations of these features, which could potentially have critical safety implications. Being able to measure or infer driver's understanding of the technology on their vehicles–and by extension, measure the efficacy of different training and education efforts–is an important area for further development.

This technical report summarizes a study appraising different classes of outcome measures and examining in an experimental setting how different measures map onto the understanding of advanced driver assistance systems. The results should be informative to researchers, the automobile industry, and government entities.

This report is a product of a cooperative research program between the AAA Foundation for Traffic Safety and the SAFER-SIM University Transportation Center.

C. Y. David Yang, Ph.D.

President and Executive Director
AAA Foundation for Traffic Safety

Dawn Marshall

SAFER-SIM Director
National Advanced Driving Simulator

The University of Iowa

About the Sponsors

AAA Foundation for Traffic Safety
607 14th Street, NW, Suite 701
Washington, D.C. 20005
202-638-5944
www.aaafoundation.org

Founded in 1947, the AAA Foundation for Traffic Safety in Washington, D.C., is a nonprofit, publicly supported charitable research and educational organization dedicated to saving lives by preventing traffic crashes and reducing injuries when crashes occur. Funding for this report was provided by voluntary contributions from AAA/CAA and their affiliated motor clubs, individual members, AAA-affiliated insurance companies, and other organizations or sources.

SAFER-SIM University Transportation Center
Federal Grant No: 69A3551747131

SAFER-SIM is comprised of a multidisciplinary, synergistic team of researchers in human factors, engineering, computer science, and psychology who will use innovative simulation approaches ranging from microsimulation to human-in-the-loop simulation to promote safety. SAFER-SIM sponsors research and outreach activities in STEM areas, and aids workforce development efforts in transportation safety.

The universities comprising SAFER-SIM study how road users, roadway infrastructure, and new vehicle technologies interact and interface with each other using microsimulation and state-of-the-art driving, bicycling, and pedestrian simulators.

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The report is funded, partially or entirely, by a grant from the U.S. Department of Transportation's University Transportation Centers Program. However, the U.S. Government assumes no liability for the contents or use thereof.

Table of Contents

Executive Summary

Advanced vehicle technologies have the potential to improve safety and convenience for drivers; however, in order to realize these benefits, it is important that drivers use the systems appropriately. Past research has documented gaps in drivers' understanding of how new technology works and when it should be used. It follows that the interplay between a driver's knowledge, performance and safety outcomes, and approaches for training and education have become topics of interest in the research and stakeholder communities. One area needing further exploration is the relationship between a driver's knowledge of the system and a wide variety of performance and safety outcome measures.

The main objective of this study was to identify and appraise outcome measures that could be used to assess or infer a driver's knowledge of advanced vehicle technologies and, by extension, the effectiveness of training and education on these same technologies. The study focused on drivers' knowledge of adaptive cruise control (ACC), a technology that is currently deployed and widely available. The study was comprised of two parts: (a) a review, enumeration, and appraisal of measures of performance, safety, and behavior that potentially bear on knowledge of advanced vehicle technology, and (b) an experimental study to explore and validate outcome measures that can be implemented in research or other settings to measure drivers' knowledge and understanding of advanced vehicle systems.

For Part 1, a gathering and cataloging of behavioral, performance, and safety measures was carried out through complementary tasks, including a scan of relevant literature and engineering standards. Outcomes from this scan were augmented with information gleaned from several available datasets. Additionally, a workshop was convened with subject matter experts to discuss a variety of topics related to education, training, and measurement. Based on the combined tasks, the universe of measures was refined and consolidated into a matrix of outcome measures, which documented different properties of the measures along with an appraisal of the advantages and disadvantages relative to their implementation in the context of driver's knowledge of vehicle technology. The main takeaway from Part 1 is that the measures vary greatly in terms of their implementation and complexity and each category or measure has advantages and disadvantages with respect to its relevance to system knowledge and to training and education. The use of combinations or clusters of outcome measures would offer stronger and more stable insights into a driver's underlying knowledge, accounting for some of the inherent shortcomings of just a single measure or class of measures.

For Part 2, a driving simulator study was executed at two different sites using a common approach, scenarios, and measures. The purpose was to identify and validate outcome measures that can be implemented to measure drivers' knowledge and understanding of advanced vehicle systems. The selection and inclusion of outcome measures was guided by Part 1, as well as by practical constraints of the experimental set up. Sixty-five participants completed the study across both sites, using an ACC system in a variety of driving conditions, including takeover scenarios, non-takeover scenarios, and during system interactions (e.g., changing settings), among others.

A machine learning approach was used to assess the effectiveness of different outcome variables in predicting driver knowledge of vehicle technology. The models highlighted the variables most strongly related to a driver's knowledge of ADAS and how these varied by measurement window; in particular, measurement windows or epochs that involved system interactions, control takeovers, or routine situations where the system is operating showed stronger outcomes in the modeling exercise.

Measures of eye glance behavior also featured prominently in all of the relevant models. This outcome corroborated some of the perspectives gleaned from Part 1, noting the advantages of eye glance metrics in mapping onto a driver's roles and responsibilities while using the technology. With respect to measures of vehicle control, some constraints were noted in Part 1 regarding their use as these measures are often impacted by many other factors. This too was evident from the experimental study, with vehicle control and related safety measures not showing as prominent components of the models in most cases. The noteworthy exception was in cases where active vehicle control was necessary (i.e., takeover scenarios).

Ultimately, employing a cluster of measures is advised in order to account for some of the limitations associated with individual measures. However, in cases where there are constraints in what types of information can be gathered or weighed, researchers and other stakeholders should do the following:

- Prioritize measures of eye glance behavior to corroborate a driver's knowledge of the vehicle system. Vehicle control and safety measures are more effectively applied in specific edge-case or takeover situations.
- Establish measurement windows around system interactions (e.g., changing system settings), system takeovers or disengagements, and normal system operation where the system is activated.
- Leverage subjective measures such as confidence or technology acceptance whenever possible, which are strong predictors of knowledge.
- Incorporate or consider information about the driver, such as driving experience and demographics.

Introduction

Advanced vehicle technologies—from Advanced Driver Assistance Systems (ADAS) to different forms of vehicle automation—are becoming more prominent in the passenger vehicle fleet. These technologies are designed to improve convenience and safety for drivers (e.g., Bengler et al, 2014; Fisher et al., 2020); however, to reap these benefits, it is important that drivers use the systems appropriately. That is, systems that take on more of the driving task, such as adaptive cruise control (ACC) and lane centering assist (LCA), should only be used in appropriate situations and environments, and with clear and accurate knowledge about what the systems can and cannot do. Thus, there is significant interest and importance in examining the mapping of advanced vehicle technologies on drivers' understanding of the technology. This is especially important in situations that are outside or nearing the limits of a system's operational design domain (ODD), which define the conditions under which a system is safe to operate.

There are multiple contemporary research efforts being undertaken to better understand the benefits, risks, and human factors issues arising from these technologies (e.g., Yang et al., 2023; Naumann et al., 2023). Several studies have examined the correspondence between a driver's knowledge of advanced technologies and the manner in which they interact with the system or their relative safety when using the technology. For example, Gaspar et al., (2020) documented performance deficits in certain scenarios for drivers that had a poorer mental model of ACC, which is a reflection of the driver's knowledge or understanding of the system, compared to drivers that had a more complete understanding of the system. The deficits in drivers' understanding of technology observed by Gaspar et al. (2020) and by others (e.g., McDonald et al., 2018; Lenneman et al., 2020; Forster et al., 2019) have helped underscore the need to examine approaches to mitigate these gaps. Training and education are one such avenue towards accelerating users' understanding of vehicle technologies and, ideally by extension, increasing their safe and appropriate use.

A number of studies have examined how different types of information or training approaches can shape system knowledge (e.g., DeGuzman & Donmez, 2022; Forster et al., 2019; 2020; Singer & Jenness, 2020). Such evaluations have employed a wide variety of approaches as well as a variety of metrics, including direct knowledge tests, driver behaviors and performance, safety and driving outcomes, a variety of subjective and physiological measures, and other outcome measures. What is less clear is the relationship between this array of measures and knowledge of vehicle technology. For example, what measures change (ameliorate or degrade) with different levels of system knowledge? Identification and evaluation of measures that can help to bridge the gap between knowledge and a variety of performance and safety outcomes will be essential for understanding the effectiveness of training or education interventions.

Objective

The main objective of this study was to identify and appraise outcome measures that could be used to assess or infer knowledge of advanced vehicle technologies and, by extension, the effectiveness of training and education on these same technologies. The study focused on drivers' knowledge of ACC, a technology that is currently deployed and widely available. Where relevant, the implications of study outcomes for other forms of ADAS and for higher levels of automation are discussed.

The study was conducted in two parts, which are described in the following sections:

- **Part 1: Cataloging Measures of Behavior, Performance, and Safety**—Enumerate and catalog measures of behavior, performance, and safety in the driving domain that potentially bear on knowledge of advanced vehicle technology. This effort was grounded in a review of the scientific literature, engineering standards, and existing datasets, and through solicitation of input from domain experts.
- **Part 2: Experimental Study—**Identify and validate outcome measures that can be implemented in research to measure drivers' knowledge and understanding of advanced vehicle systems using a multi-site experimental approach. A driving simulator study was executed at two different sites using a common approach, scenarios, and measures.

Part 1: Cataloging Measures of Behavior, Performance, and Safety

The gathering and cataloging of behavioral, performance, and safety measures was carried out through three complementary and concurrent approaches. Researchers conducted a scan of relevant literature as well as engineering standards. Outcomes from this scan were augmented with information gleaned from several datasets gathered in recent and related studies from the University of Massachusetts–Amherst Human Performance Laboratory (HPL) and partner universities at the SAFER-SIM University Transportation Center (UTC). Additionally, a workshop was convened with several subject matter experts to discuss a variety of topics related to education, training, and measurement.

Based on the combined tasks, the universe of measures was refined and consolidated into a matrix of outcome measures. The matrix documented different properties of the measures along with an appraisal of the advantages and disadvantages relative to their implementation in the context of knowledge of vehicle technology. A subset of subject matter experts from the workshop provided additional feedback on the matrix.

Literature, Standards, and Dataset Review

Literature searches were conducted using the Transportation Research International Documentation (TRID) database, as well as Google Scholar. Articles appearing in journals, conference proceedings, and technical reports in English were considered for inclusion. Search terms included various combinations of keywords, such as driver or driving, training or education, automation, ADAS, adaptive cruise control, or lane keeping assist, and other like terms.

The initial search yielded an expansive list of articles of over 5000 articles. The titles and abstracts of the articles were then screened by the research team for relevance. A final set of 49 articles was compiled and downloaded for further review. Study details, along with features and characteristics of the measures were extracted into a table for further processing.

Standards documents were also identified in the searches. Additionally, known standards from sources such as the Society for Automotive Engineers (SAE; e.g., J2944, J2396) and the International Organization for Standardization (ISO) were also reviewed.

Lastly, outcome variables gathered in the datasets from six recent studies by the University of Massachusetts–Amherst HPL and some universities in the SAFER-SIM UTC were reviewed for any additional data elements that could be considered in the data catalog.

A comprehensive list of over 120 variables was drawn from the collective sources. The measures included performance and driving outcomes, physiological indices, subjective ratings, as well as measures of driver knowledge. Many variables were similar or related and so the research team went through an iterative process to group or categorize measures into a more finite matrix (described below).

Subject Matter Expert Workshop

A 3-hour in-person workshop was convened on January 8, 2023, to solicit input from a group of subject matter experts concerning driver knowledge, education, and training in the context of advanced vehicle technologies, as well as regarding issues surrounding measurement. The workshop included two experts from academia, one from industry, and one from the government (as well as the research team). Following the welcome and a general discussion, the group entertained a round table discussion on current and future measures to evaluate the effectiveness of consumer education and the assessment of different measures.

The experts volunteered a number of both specific and broad categories of measures that could be useful in this space. These included the following:

- Measures of system interactions, including but not limited to observations as to whether drivers are using the systems when they should. This could also encompass basic information about system usage (% use, frequency of use), whether drivers can accurately turn on the system, etc.
- User experience (UX)-type measures, including satisfaction and quality of interactions. Such measures can be examined in light of individual differences across user groups.
- Driver knowledge itself, as a direct measure, as an important outcome.
- Eye glance behavior, both during initial system use but also as drivers become more familiar with the system. Similarly, head position might be a useful proxy for some elements of glance behavior and would be easier to implement.
- Strategy as an outcome measure, or why drivers choose or choose not to use the system.
- Pre- and post-event behavioral measures, regardless of whether the event was positive (e.g., system acted as it was supposed to) or negative (e.g., system failed to do something it was supposed to).
- Various measures gathered through the course of training might be informative, such as click throughs on training content or other indices of engagement.
- Composite measures might be feasible to define different driver profiles or levels of at-risk drivers, though what measures compose such a composite remain open.

Though not an outcome measure per se, the inclusion of contextual information was considered to be highly beneficial in support of other performance and safety measures. The experts also flagged some important corollary issues that are worthy of consideration.

- Understanding what is intended by education and training (i.e., how to operationalize these terms). While people often use the terms interchangeably, they carry different connotations with different stakeholders. For example, original equipment manufacturers (OEMs) tend to not be in favor of using training from a liability standpoint. They also noted that there could be challenges dissociating effects due to training and education from those due to exposure alone.
- In some cases, drivers can reap the benefits of technology without necessarily understanding it (e.g., where design is also a factor); that is, appropriate system use (and positive downstream safety outcomes) might be realized even if there is a potential mismatch in knowledge.
- Importance of how system understanding evolves with experience with the system and based on the types of feedback drivers receive from the system. Along these lines, there should be a balanced consideration of immediate outcomes versus lagging or downstream indicators.

- There are some circumstances where knowledge could lead to worse outcomes, say in the case of deliberate circumventing of known system limits (e.g., knowing how to “game” a driver monitoring system). Similarly, for some measures, it might be difficult to discern poor performance from willful disregard or poor comprehension.
- Though not without challenges, certain outcome measures could be tracked over the long term; in that sense, vehicles could host a personalized history of the driver along with driving and event data. This could capture trends and changes in behavior and safety over time, including adaptation.
- It is important to not only look at candidate measures, but to appraise how feasible it is to incorporate them into production vehicles.
- The transfer of training from one system or vehicle type to another is generally not well understood in this space.
- There is a need to consider who is using the data and for what purpose; for example, researchers, driving instructors, DMVs, advocates, etc.

The information gleaned from the experts helped to reinforce a number of the outcomes identified in the literature, standards, and dataset reviews. It also helped to expose some new dimensions of potential importance, including more outcomes or derivations related to system interactions and UX measures, as well as a number of different lenses through which the outcome measures should be considered.

Behavioral, Performance, and Safety Measures

As noted above, a list of over 120 variables was drawn from the various sources. The measures included performance and driving outcomes, physiological indices, and subjective ratings, as well as measures of driver knowledge. Based on input from the expert workshop, other indices related to system interactions and UX were incorporated as well as a number of dimensions along which each measure was weighed.

Many variables were similar or related. In order to simplify the universe of measurements and related discussion, the research team engaged an interactive process wherein they discussed and distilled outcomes from the main matrix into broad categories. Lastly, the subject matter experts from the workshop were solicited for some additional input regarding the matrix, leading to some additional changes. The final matrix is shown below in focused form (with the full, expanded version available in spreadsheet format here).

The matrix delineates nine categories of outcome measures:

- ***Control transitions***: measures related to the transition of vehicle control or partial control from the automation to the driver or vice versa
- ***System interactions***: measures related to system use (e.g., how and when the system is used, when system settings are adjusted)

- ***Vehicle control***: measures related to the lateral and longitudinal control of the vehicle
- ***Safety events***: critical incidents, including crashes or near misses
- ***System understanding***: measures of knowledge or confidence in knowledge
- ***Attitudes and perceptions***: acceptance, trust, and behavioral intentions towards automation, among other subjective measures
- ***Workload***: measures related to the level of mental or physical effort required or stress experienced when interacting with the system
- ***Usability***: measures of user experiences and satisfaction
- ***Attention***: measures of attention allocation and engagement with different tasks

Under each category, a number of sample or representative measures are listed. In some cases, the sample measures themselves represent an array of related measures. For example, measures of lane position can be reflected by many specific measures, such as standard deviation of lane position, root mean square error, and the like.

For each category or sample measure, the matrix further details the complexity of the measure (how challenging it would be to instrument and gather these measures), the context or under what conditions the measure might be assessed, and whether the measure is specific to a unique circumstance or use-case scenario. Advantages and disadvantages of the measure are provided in relation to its utility for appraising knowledge or the impact of education and training. Lastly, the full matrix (link) identified the timing of the measure relative to driving; whether the measure maps onto skills, rules, or knowledge (SRK; based on Rasmussen's (1983) framework); if training could conceivably target the measures; and the expected impact of training on the outcome.

Category	Measure	Context	Complexity	Pros	Cons
Control Transitions	Accuracy or success in situations where drivers takeover control from automation (e.g., successful takeover/deactivation counts or rates)	Situations involving a state change, where the automation is engaged or disengaged	Complex, depending on necessary sensors and operational setting as well as clear criteria to define success.	- Used extensively in automation research, allowing for comparison - High face validity with respect to training; i.e., direct mapping to certain functions provided by automation (e.g. ADAS and ACC functions) - Shown in past studies to be sensitive to differences in level of understanding in some use cases (Gaspar et al., 2020) - Can be monitored indefinitely (depending on implementation)	- Impacted by other factors outside of education/training (e.g., experience, vehicle automation technologies, implementation) - Past use in other studies is not necessarily indicative of a good measure - Unknown if sensitive to differences in training and in what situations and on what time horizon (i.e., how proximal to training)
	Quality of a control takeover (e.g., response time, smoothness of control, duration)	Situations where control returns to driver (e.g., system edge cases)	Moderate, depending on necessary sensors and operational setting		
	Driver-initiated disengagement of system in situations where it can operate safely (e.g., frequency)	Situations involving a state change, where the automation is disengaged	Simple, depending on clear mapping of system capabilities and conditions; however, could be complex, if details of the context are required to define 'necessity'	- Errors can be related to Skills-Rules-Knowledge - High face validity with respect to training - Measuring errors is a standard assessment approach - Expert (human) assessments possible without 'processing' data -This, along with system use, might be an index of the value of the technology to the user	- Not frequently used in safety literature - Relates to system performance as opposed to safety outcomes (though error-prone systems are likely to correlate with safety) - Could be challenging to operationalize what is considered 'unnecessary'
	Efficiency or success in handing off control from driver to system (appropriate/timing of system engagements; failure to engage properly)	Situations involving a state change, where the automation is engaged	Simple, depending on clear understanding of system capabilities in relation to current conditions	- Quality of handoffs can be related to Skills-Rules-Knowledge - High face validity with respect to training	- Requires precise/careful definition and criterion - Limited by availability of systems (e.g., ODD)
System Interactions	Accuracy of system interactions, other than control transitions (e.g., error rate)	Situations involving a change to system settings	Simple, with knowledge of current and desired system state or setting	- Errors can be related to Skills-Rules-Knowledge - High face validity with respect to training - Measuring errors is a standard assessment approach - Expert (human) assessments possible without 'processing' data	- Not frequently used in safety literature - Relates to system performance as opposed to safety outcomes (though error-prone systems are likely to correlate with safety) - Measures such as error rate alone might not capture the impact of different types of errors on performance or safety
	Driver position and posture (e.g., frequency/rate of hands/feet hovering near controls, body position)	Normal system operation and/or state transitions	Simple, provided clear criteria are established; however, can be labor intensive depending on coding approach	- Used extensively in transportation safety research, allowing for comparison - Can be monitored indefinitely (depending on implementation)	- Impacted by other factors outside of education/training (e.g., driver characteristics, fatigue, trust in technology, driving duration, traffic situation) - Unknown if sensitive to differences in training - Difficult to operationalize and code
	System use (e.g., % time system engaged, frequency)	All driving conditions (manual and system)	Simple, depending on the denominator (i.e., total trip time versus time system is available for use)	- Use can be related to Skills-Rules-Knowledge - Moderate face validity with respect to training -This, along with system disengagements, might be an index of the value of the technology to the user - Could be parsed according to safe/appropriate use	- Not frequently used in literature - Can be dependent on factors outside of education/training (e.g., trust, experience, whether automation system is available to be used [system ODD]) - Unknown if sensitive to differences in training

Category	Measure	Context	Complexity	Pros	Cons
Vehicle Control	Speed and speed variance, acceleration (e.g., maximum, SD)	Normal system operation and/or state transitions	Simple, depending on available data	- Used extensively in transportation safety research, allowing for comparison - Can be monitored indefinitely (depending on implementation) - Continuous measurement in normal conditions can offer insight into routine system use - Edge-cases shed insight into handling and resolution of critical events -Many of these measures can be leading indicators of safety	- Impacted by other factors outside of education/training (e.g., speed limit, road type, traffic level, driving style, experience) - For certain measures, clarity regarding who is controlling or accountable for a given outcome might be needed (i.e., driver or the automation) - Low face, construct, and content validity with respect to education/training (depending on the implementation) - Unknown if sensitive to differences in training (e.g., criterion validity)
	Headway (e.g., time/distance, min. time-to-collision, SD headway)				
	Lane position (e.g., SD/RMSE of lateral position, max deviation)				
	Steering (e.g., SD steering position, max. steering, steering entropy)				
Safety Events	Crashes (e.g., presence, rates)	Normal system operation and/or state transitions	Simple, depending on sensors and available data	- Used extensively in safety research, allowing for comparison - Critical metrics in influencing policy, practice, etc. - Can be monitored indefinitely (depending on implementation) - Continuous event monitoring can offer insight into routine system use and edge-case situations - Can and should be examined along with measures of transitions of control, noted above	- Impacted by other factors outside of education/training (e.g., experience, skills, traffic situation) - Low face, construct, and content validity relative to training (depending on the implementation) - Unknown if sensitive to differences in training - Such events range from rare to extremely rare; gathering sufficient data requires an effort to scale
	Near misses		Complex; requires clear criteria and is difficult to capture		
	Lane departure (e.g., presence, rates)		Simple, depending on sensors and available data		
System Understanding	Confidence in knowledge	Questions can relate to general use or specific use cases	Simple	- Focused exclusively on driving automation - Strong face and construct validity with respect to education/training - Potential for good content and criterion validity (i.e., can be standard against which other measures are examined) - Administration can be recurring - Assessment is immediate - Can be done in situ with real time probes - Examining confidence and knowledge together can add another dimension to the data (given possible co-variation)	- Subjective assessment may not correspond to actual behavior or specific safety benefits - No common or standard survey tool (i.e., threats to content and criterion validity) - Surveys can be long - Single assessment does not capture evolution of mental model over time - Depending on how administered, not sensitive to knowledge gained in training versus a priori knowledge - Currently many different survey instruments exist (which could be a pro as well)
	Knowledge assessment				
Attitudes and Perceptions	Acceptance of technology (e.g., intentions to use/planned adoption, affinity for technology, perceived usefulness)	Questions can relate to general use or specific use cases	Simple	- Acceptance of, intentions to use, and perceived utility are likely related to knowledge of technology - Assessment is immediate - Some constructs grounded in Theory of Planned Behavior	- Impacted by other factors outside of education/training (e.g., age, experience, tech savviness) - Low face, construct, and content validity relative to training (i.e., not to confuse with knowledge itself)
	Trust or confidence in technology	Questions can relate to general use or specific use cases	Simple	- Trust and confidence in technology are likely connected to knowledge of technology - Frequently used/cited in driving automation research - Assessment is immediate	- Impacted by other factors outside of education/training (e.g., age, experience, tech savviness) - Low face, construct, and content validity relative to training (i.e., not to confuse with knowledge itself) - Trust is multifaceted; different metrics and approaches exist
	Attitudes towards training (e.g., preferred learning method; evaluation of training)	N/A	Simple	- Adopting preferred or highly rated approach might increase likelihood of training, though does not bear on knowledge directly	- Impacted by other factors outside of education/training (e.g., age, experience, learning style) - Unclear association with knowledge gleaned through training/education

Category	Measure	Context	Complexity	Pros	Cons
Workload	Subjective ratings of workload or stress (e.g., Rating Scale of Mental Effort, NASA-TLX)	Normal system operation and/or state transitions	Simple	- Standardized metric - Easy to administer - Index of driver capacity - Assessment is immediate - Frequently used in transportation safety/human factors research - Can be particularly helpful to measure drivers' workload following critical events	- Subject to significant noise and potential bias - Impacted by many extraneous situational and individual factors - Challenging to measure in situ in operational settings
	Physiological measures of workload and arousal (e.g., cardiac metrics—HR variability, root mean square of successive differences (RMSSD); pupil metrics—pupil diameter, blink frequency; galvanic skin response)	Normal system operation and/or state transitions	Moderate, requiring specialized sensors and synchronization to other sources of data or to the driving environment	- Standard and objective measurement - Extensive literature - Can be particularly helpful to measure drivers' workload around the critical events	- Measurement burden/intrusive to measure - Subject to significant noise - Impacted by many extraneous situational and individual factors
	Brain Activity (e.g., electroencephalography (EEG); cerebral hemodynamics/oxygenated hemoglobin (Oxy-Hb))	Normal system operation and/or state transitions	Complex, requiring highly specialized and intrusive sensors and synchronization to other data sources or the driving environment	- Relatively novel in driving studies, but extensive research in other domains - Can be particularly helpful to measure drivers' workload around the critical events	- Measurement burden/intrusive to measure - Subject to significant noise and can be difficult to interpret - Highly specialized, requiring expertise - Impacted by many extraneous situational and individual factors
Usability	Usability survey or scale (e.g., System Usability Scale (SUS); User Experience Questionnaire, (UEQ))	Questions can relate to general use or specific use cases	Simple	- Used extensively in computer science/HCI research - Standardized scales available	- Impacted by other factors, such as system/human-centric design - Not necessarily influenced by education/training
Attention	Attention allocation (e.g., % distribution of glances to areas of interest (AOIs); on-road/off-road glances, total eyes-off-road time, glance counts to AOIs)	Normal system operation and/or state transitions	Moderate, requiring specialized equipment and capabilities to post-process data, as well as means to define critical regions of interest using global coordinates (for AOIs)	- Used extensively in transportation safety research, allowing for comparison - Maps onto driver responsibility to be attentive while using some types of automation - Can be monitored indefinitely (if implemented) - For some use cases, head movements might be reasonable proxy for eye movements.	- Impacted by other factors outside of education/training (e.g., experience, situational factors, driver state, presence of secondary tasks) - Measurement burden, difficulties in tracking some individuals - Modest face, construct validity with respect to education/training (depending on implementation); low content validity - Unknown if sensitive to differences in training - Requires precise definitions of AOIs (i.e., glance durations to "what"—whether forward roadway or elsewhere)
	Glance metrics (e.g., mean duration, SD, max. duration, proportion of long glances, frequency, glance rates, time to return to roadway)				
	Engagement in non-driving related tasks (NDRT) (e.g., frequency, duration)	Normal system operation and/or state transitions	Moderate, depending on desired precision of measures (e.g., task = yes or no versus time on task)	- Widely used in studies of vehicle automation; - NDRT engagement may serve as a surrogate for knowledge/trust.	- Impacted by other factors outside of education/training (e.g., age, experience, traffic conditions) - Need to consider the nature of the NDRT as tasks are not equivalent - At higher levels of automation, NDRT is permitted
	Hazard detection and anticipation (% success)	Normal system operation and/or state transitions	Complex, requiring information about hazards as well as driver's detection or anticipation of them (captured overtly through eye glances or covertly through behavioral proxies)	- Widely studied in safety literature (albeit not in 'technology' domain) - Hazard anticipation in certain scenarios could map onto knowledge/mental models	- Impacted by other factors outside of education/training (e.g., age, experience, traffic conditions) - Connection between these metrics and education/training on vehicle technology is unclear (though good support for hazard anticipation training itself)

Discussion

The purpose of this exercise was to enumerate and catalog measures of performance, safety, and behavior in the driving domain that could potentially bear on driver knowledge of advanced vehicle technology and, by extension, to the effectiveness of training and education approaches. A review of the scientific literature, engineering standards, and existing datasets was conducted, extracting numerous outcome measures of varying types. Additionally, input from domain experts was used to establish and then refine the list and appraisal of different measures.

The initial scan of the different sources yielded an expansive list of over 120 variables; however, removal or combination of redundant or very similar measures led to nine broad categories of measures in the matrix. The main takeaway from the matrix is that the measures vary greatly in terms of their implementation and complexity and each category or measure has advantages and disadvantages with respect to its relevance to system knowledge and to training and education. Moreover, some outcome measures can be targeted directly through training (e.g., teaching drivers how to turn on the system or when to take over control), while other measures are less direct. That said, in many cases, the relationship between a given outcome measure and driver knowledge has not been empirically established. Moreover, safe and appropriate system use, rather than driver knowledge per se, might be the appropriate target for measurement. Though not exhaustive relative to the full matrix (link), some highlights for each measurement category are discussed below.

Given its strong face and construct validity, direct measurements or assessments of driver knowledge might be considered a gold standard. However, it is recognized that measuring driver knowledge directly might not always be feasible, nor can it necessarily be measured over time easily. Moreover, accurate knowledge of a system might not necessarily or always translate to accurate or appropriate system use (e.g., cases where drivers knowingly push the system beyond its limitations).

Many of the measures related to control transitions and system interactions were thought to have high face validity as they often reflect appropriate and safe use of the systems. Such knowledge could be directly targeted through training. However, outside of controlled laboratory or experimental settings where these measures have often been employed, these could be complex depending on the setting/situation. There is some evidence that these classes of measures are sensitive to different levels of system understanding (e.g., Gaspar et al., 2020).

Vehicle control and safety measures, such as speed, lane keeping, and crashes, are commonly used in traffic safety research and can be readily monitored continuously in all driving contexts. One major drawback, however, is that these measures are also impacted by other factors besides driver knowledge, such as environmental or situational factors. Crashes and near misses, including lane departures, are also very rare events in the real world and so could be difficult to leverage in an assessment of knowledge of vehicle technology.

Many measures of attitudes and perceptions are available and have been deployed in past studies, such as trust in and adoption of technology and perceived useability of the systems. While these can be easily administered, they often deal with constructs that are related to driver knowledge of technology. Similar to the performance outcomes noted above, these measures can also be impacted by other factors and have relatively low face and construct validity relative to system understanding.

Measures of workload vary in the degree of challenge in implementation and can offer important insight into the driver's state or in their response to different situations (with or without the vehicle systems activated or during control transitions). That said, workload responses are, like many other measures, also highly susceptible to noise and other extraneous sources.

Measures of driver's attention, such as eye glance behavior and scanning also vary in terms of the complexity of implementation, but benefit from extensive use in past research, as well as innovative approaches to make real world use more practicable. While attentional allocation can be impacted by other extraneous factors, one discernible benefit is that driver attention can be mapped directly to key knowledge or training content regarding driver responsibilities while using certain vehicle technology. For example, measures of attention can offer an index of drivers' vigilance and alertness while using the systems (and not engaging in non-driving related tasks).

Outcomes from the matrix and from feedback from the subject matter experts underscore that the use of combinations or clusters of outcome measures would offer a stronger and more stable insight into a driver's underlying knowledge. Such an approach would help account for some of the inherent shortcomings of just a single measure or class of measures. This would also help guard against potential homeostasis for some measures, where increased knowledge or training leads to improvements on some outcomes, but could also lead to more use of automation in more challenging situations.

Building on the collection and classification of outcome measures, as well as their appraisal through the matrix, Part 2 of the study was intended to provide some additional insight into the efficacy of different measurement types under different use cases.

Part 2: Experimental Study

The purpose of the experimental study was to identify and validate outcome measures that can be implemented to measure drivers' knowledge and understanding of advanced vehicle systems. The selection and inclusion of measures was guided by Part 1, as well as by practical constraints of the experimental set up. Direct assessment of system understanding (via knowledge survey) was chosen as the "gold standard" for this effort, against which other metrics were evaluated.

In order to gather a larger and more diverse sample of drivers, a multi-site experimental approach was adopted. A driving simulator study was executed at two different sites—UMass in Amherst, MA, and AAAFTS in Washington, DC—using a common platform, approach, scenarios, and measures. Given the aim of the study, to map differences in knowledge onto different performance outcomes, it was important to include participants with a wide range of knowledge of ACC. This was achieved through participant screening, as well as selective training of individuals at the outset of the experimental session.

The human subjects experiment was overseen by two Institutional Review Boards (IRB), one for each site. The protocol was approved by the University of Massachusetts IRB (Study #4718) and by Advarra IRB (Pro00073933) and performed in accordance with the ethical standards as described in the 1964 Declaration of Helsinki.

Method

Participants

Both sites recruited participants from their respective communities via flyers, ad postings on Craigslist, and word-of-mouth. To be eligible, participants were between 18 and 65 years old, had a valid driver's license, and were not susceptible to motion sickness. Characteristics for the sample of drivers are provided in Table 1.

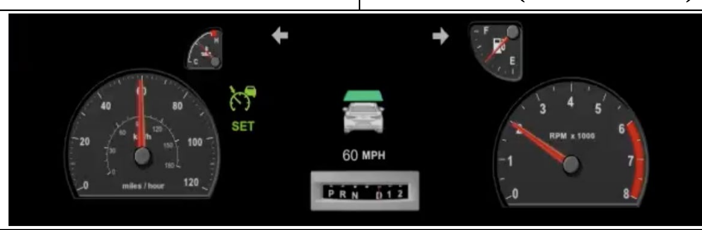
Table 1. Sample characteristics for the two sites.

	UMass	AAAFTS	Overall
Sample, N	**41**	**24**	**65**
% Female (N)	32.3% (N = 10)	67% (N = 16)	40% (N = 26)
Mean Age (SD)	25.5 (6.0)	33.8 (9.9)	28.6 (8.6)
Mean Driving Experience, Years of Licensure (SD)	4.5 (5.5)	15.9 (8.9)	8.4 (8.6)
Weekly Miles, N			
<50	28	11	39
50–100	8	10	18
>100	5	3	8
Experience with ACC (%)	49%	17%	37%

Apparatus

Driving Simulator. Both sites used a fixed-base Realtime Technologies Inc. (RTI; Ann Arbor, MI) driving simulator with SimCreator version 3.8 Build 10.9.19 software. Table 2 provides a side-by-side comparison of the hardware features for each site.

Table 2. Driving simulator characteristics for both sites.

Hardware	UMass	AAAFTS
Configuration	 Ford Fusion Full Vehicle	 Racing Simulator Cockpit
Number of Screens	6 total: 5 front screens + 1 rear screen	3 monitors
Field of View	330 degrees	100 degrees
Instrument Cluster	In-cab display (Altia software)	Simulated on front screen (Altia software)
ACC Visual Interface		
ACC Controls	OEM wheel with physical buttons 	Logitech G27 USB racing wheel with physical buttons and controls on gear box 
Center Console	Touchscreen	Physical controls
Speakers	5.1 channel audio system + 2 internal (vehicle) speakers	Logitech Z313 speaker system
Screen Resolution	Front projectors: 1920 x 1200 Rear projector: 1400 x 1050	Monitors: 5760 x 1080

Eye Tracking and Video Monitoring. At the AAAFTS site, a Dikablis Model 3 eye tracker and D-lab software (Manching, Germany) were used to capture and analyze eye tracking data. The system optics were mounted in a lightweight pair of glasses worn by participants. The UMass site employed a SmartEye tracking system (v.9.1, Gothenburg, Sweden) with four SmartEye Pro cameras mounted on the dashboard and center console. While the two systems differed in terms of their implementation (e.g., head-mounted versus remote), both systems were able to generate comparable outcomes according to the lists of dependent measures noted below.

At the UMass site, participants were recorded using the video capture and review system, Sim Observer (RTI, Ann Arbor, MI). Two cameras recorded the participant's hand and foot movements, as well as the forward view and instrument cluster.

Driving Environments and Scenarios. Three experimental drives were developed within a rural, lightly industrialized environment (Figure 1). The road was a four-lane road with two lanes in both directions; in some sections, the road was divided and had a wide right-hand shoulder whereas in other sections the road was undivided and had a narrow shoulder. The posted speed limit was 60 or 65 mph, depending on the location, and traffic in the oncoming lane was sparse. In one drive (Drive 2), the ambient lighting was reduced to simulate dusk.

Figure 1. Sample image of the driving environment.

The ACC system was activated using either the steering wheel buttons (UMass site) or the center gear box (AAAFTS site). Participants could set the target speed, as well as the headway distance/spacing. The system was intended to operate effectively on all roads in the study (the scenarios described below notwithstanding). When activated, the ACC would maintain the set speed and desired vehicle spacing (see image in Table 2); drivers could adjust the speed upwards or downwards and the spacing closer or farther using physical controls.

Within each drive, participants encountered varying scenarios, probes, and cues intended to demonstrate either understanding or appropriate use of the ACC system. The scenarios, probes, and cues within each drive are described below in the order in which they took place in each drive (see Tables 3, 4, and 5). Some scenarios did not require a control transition (Non-Takeover Scenarios); that is, the event did not present a situation outside of the ACC capabilities (i.e., outside of the ODD). Other situations did exceed or encroach upon the ACC system capabilities and thus required that drivers takeover control (Takeover Scenarios; Figure 2). Each drive was counterbalanced across participants.

Table 3. Scenarios, probes, and cues encountered in Drive 1.

Type	Event (label)	Description
Non-Takeover	Pedestrian crossing (D1_S1)	The driver approaches an intersection with traffic lights (displaying green) and a pedestrian approaches the edge of the crosswalk in the driver's travel path
Probe	"What is the current set speed?" (D1_P1)	The driver is asked to provide a verbal response
Non-Takeover	Shopping area (D1_S2)	The driver encounters a shopping area on their right side with stores and several parked vehicles
Probe	"What is the current distance setting?" (D1_P2)	The driver is asked to provide a verbal response
Takeover	Straddling lead vehicle (D1_S3)	The driver encounters a lead vehicle that is swerving in and out of their travel lane
Cue	"Deactivate ACC without using your pedals" (D1_C1)	The driver is asked to deactivate ACC using "Cancel"
Cue	"Reactivate ACC at the posted speed limit" (D1_C2)	The driver is asked to reactivate ACC
Takeover	Slow-moving traffic (D1_S4)	The driver reaches a line of slow-moving vehicles on the highway
Non-Takeover	Following Lead Vehicle (D1_S5)	The driver encounters a lead vehicle traveling at the speed limit

Table 4. Scenarios, probes, and cues encountered in Drive 2.

Type	Event	Description
Takeover	On-coming vehicle (D2_S1)	The driver encounters a construction zone on the opposite side of the roadway and a vehicle moves into their travel lane from the opposite direction to pass the construction zone
Probe	“Is the current vehicle speed lower than the ACC set speed?” (D2_P1)	The driver is asked to provide a verbal response
Non-Takeover	Construction zone on opposite side (D2_S2)	The driver encounters a construction zone on the opposite size of the roadway
Probe	“Is ACC currently active?” (D2_P2)	The driver is asked to provide a verbal response
Takeover	Oversized vehicle (D2_S3)	The driver encounters an oversized vehicle in the right lane
Takeover	Poor visibility due to fog (D2_S4)	The driver encounters bad visibility conditions due to fog
Cue	“Set ACC speed to 65 mph” (D2_C1)	The driver is asked to increase the ACC set speed by 5 mph
Cue	“Set ACC speed to 60 mph” (D2_C2)	The driver is asked to set the ACC set speed back to the posted speed limit
Non-Takeover	Following Lead Vehicle (D2_S5)	The driver encounters a lead vehicle traveling at the speed limit

Table 5. Scenarios, probes, and cues encountered in Drive 3.

Type	Event	Description
Non-Takeover	Car on the right shoulder (D3_S1)	The driver encounters a car parked on the right shoulder
Probe	"What speed are you traveling at?" (D3_P1)	The driver is asked to provide a verbal response
Cue	"Set ACC to 55 mph" (D3_C1)	The driver is asked to decrease the ACC set speed by 5 mph
Cue	"Set ACC to 60 mph" (D3_C2)	The driver is asked to set the ACC set speed back to the posted speed limit
Takeover	On-coming vehicle (D3_S2)	The driver encounters an on-coming vehicle in the left lane traveling against the direction of traffic
Non-Takeover	Pedestrians on sidewalk (D3_S3)	The driver encounters pedestrians walking on the sidewalk on their right
Takeover	Bicyclist group (D3_S4)	The driver encounters a group of bicyclists in the right lane
Probe	"Is ACC currently active?" (D3_P2)	The driver is asked to provide a verbal response
Non-Takeover	Following Lead Vehicle (D3_S5)	The driver encounters a lead vehicle traveling at the speed limit

(a)
(b)

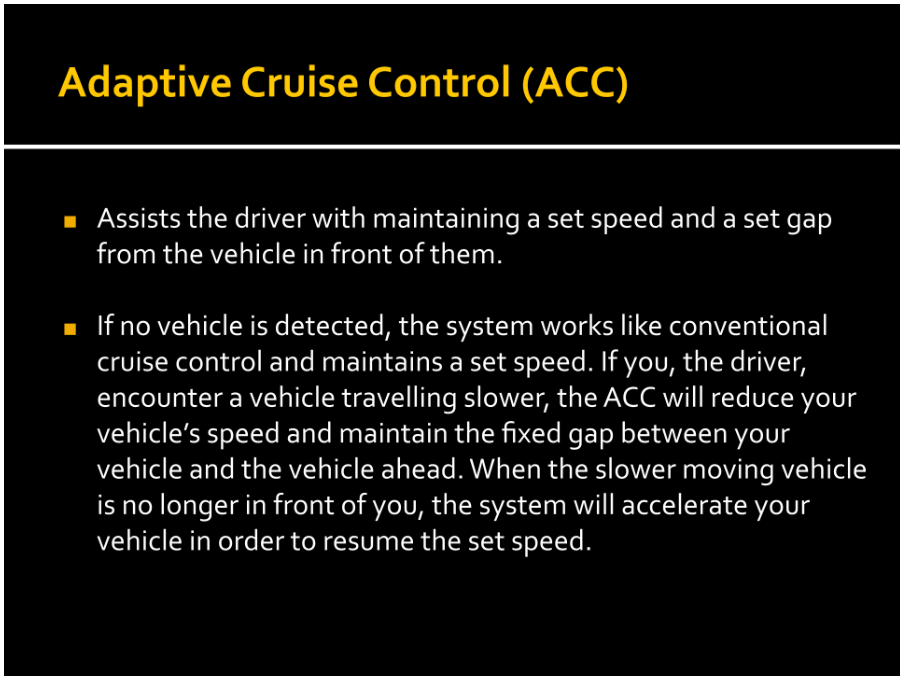
Figure 2. Examples of driving scenarios. (a) Oncoming vehicle in left lane. (b) Oversized vehicle in right lane.

Training Material. Two versions of training material were developed and these were selectively employed in order to increase the variability in knowledge across participants prior to their exposure to the experimental drives. Training content was adapted from materials used by Gaspar et al. (2020). Based on their scores on four knowledge pre-screen questions, participants were identified as being novice (score of 0 or 1) or advanced in their knowledge of ACC (score of 3 and higher). Participants that correctly answered two items were randomly assigned to one of the two conditions in order to balance groups.

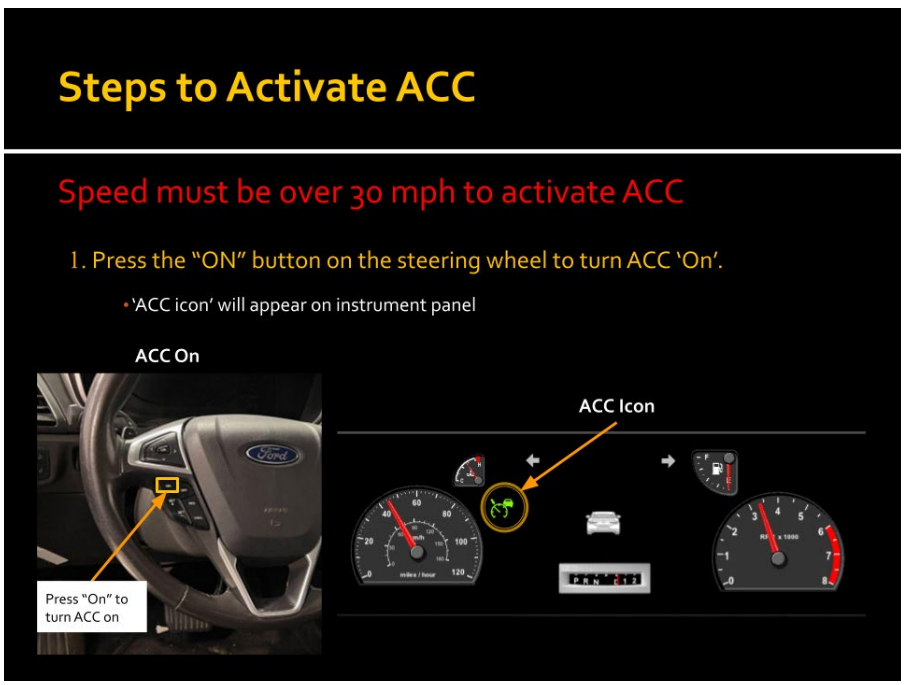
The advanced group was provided with enhanced training material, which provided information about what ACC is, how it works, conditions that must be met for use, limitations, and common scenarios where ACC might not work in addition to detailed instructions on how to activate ACC, change headway distance/spacing settings, and cancel and resume ACC in the driving simulator environment they would be using (including labeled pictures). This was intended to reaffirm and/or expand their pre-existing level of knowledge concerning ACC.

In contrast, participants in the novice group were given basic training material, which covered what ACC is and how it works (same as for the advanced group), as well as pared down instructions for how to activate, cancel, and resume the ACC system. These slides included pictures of the ACC controls they would be using, though unlike in the enhanced slides, labels explaining proper interaction with the controls were not included. This approach was intended to offer these participants only basic instructions about using the ACC in the study. Sample slides are shown in Figure 3 and the full training content is shown in Appendix A.

The novice and advanced groups were only established as interim categories to inform the selection of the appropriate training approach. For each site, approximately half of the participants fell into each category. Ultimately, participants' scores on the mental model assessment were used as the outcome measure of interest, irrespective of which training approach they encountered.


Adaptive Cruise Control (ACC)
Assists the driver with maintaining a set speed and a set gap from the vehicle in front of them.
If no vehicle is detected, the system works like conventional cruise control and maintains a set speed. If you, the driver, encounter a vehicle travelling slower, the ACC will reduce your vehicle's speed and maintain the fixed gap between your vehicle and the vehicle ahead. When the slower moving vehicle is no longer in front of you, the system will accelerate your vehicle in order to resume the set speed.

(a)


Steps to Activate ACC
Speed must be over 30 mph to activate ACC
1. Press the "ON" button on the steering wheel to turn ACC 'On'.
'ACC icon' will appear on instrument panel
ACC On
ACC Icon
Press "On" to turn ACC on

(b)

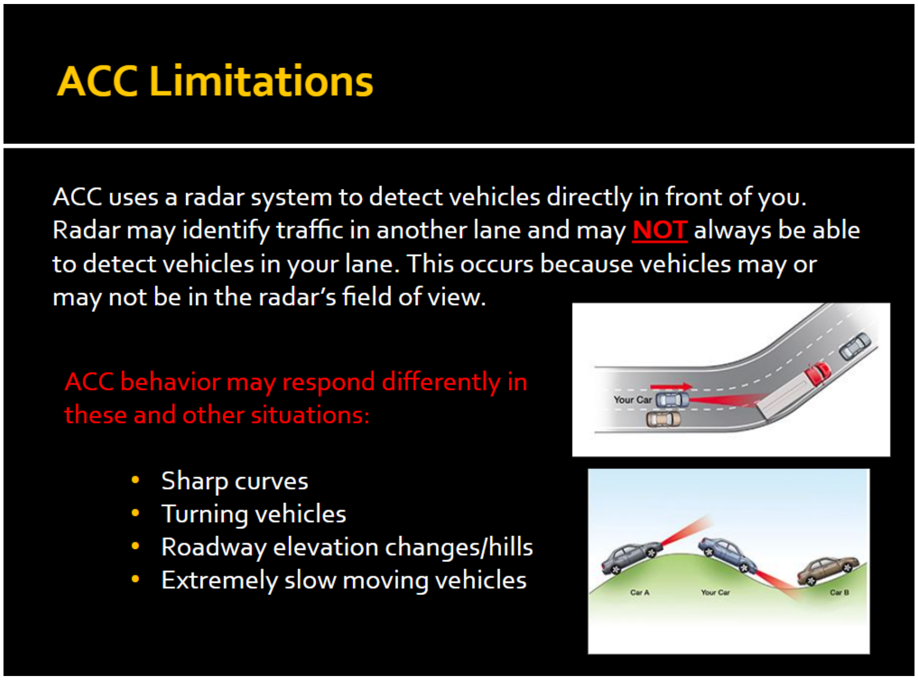
(c) 

Figure 3. Examples of training slides. (a) Basic system description (novice group, both sites). (b) Instruction on turning on system (advanced group, UMass site). (c) System limitations (advanced group, both sites).

Procedure

For each site, participants were recruited locally using posted flyers and Craigslist advertisements. In the screening survey, participants were asked to provide their contact information and responses to items that would determine their eligibility, including age, if they possessed a driver's license, four questions to assess their current level of ACC knowledge, and three questions about their potential tendency to experience simulator sickness. Eligible participants received a follow- up email from an experimenter to schedule their in-person session. Responses to the ACC knowledge questions were used to classify drivers, per the description above.

Upon arrival at the lab, participants were given an overview of the experiment and their participation and completed the informed consent process. Participants then completed a demographic survey, a trust survey (Jian et al., 2000; Appendix B), and a technology acceptance survey (Van der Laan et al., 1997; Appendix C). Following these surveys, participants reviewed one of two sets of ACC training slides, according to their knowledge group. Advanced participants were shown training slides with more robust information about the ACC system and its limitations (13 slides, Appendix A), while novice participants received a more condensed version of the slides with more basic information about the ACC system (3 slides, Appendix A). Participants were allowed to move through the training materials at their own pace, and to indicate to the experimenter when they felt comfortable with the content. After their training, participants completed the mental model assessment (Appendix D).

Participants were then introduced to the driving simulator and instructed about its configuration and assignment of standard features (turn signals, gear shift, etc.). Participants completed a practice drive lasting approximately five minutes in order to familiarize themselves with the system and to provide an opportunity for experimenters to monitor for signs of motion sickness. Participants were informed that experimenters could answer questions about the basic system, but were unable to answer any questions about the ACC feature. Following the practice drive, participants were introduced to the eye-tracking system and the system was calibrated.

Experimenters then provided participants with instructions about the three experimental drives. Participants were encouraged to drive as they would in real world conditions and use the ACC system as much as possible within these drives, but to use their best judgment if they encountered a situation they felt the ACC system could not handle. Participants were told that they would hear audio messages throughout the drives that would ask them to perform an action or to answer a prompt, and that they should try to act and respond to the best of their ability. Participants were then told that once the drives began, the experimenters were unable to answer any questions.

The order of the drives was counterbalanced across participants. After completing the drives, participants completed a second trust survey (Jian et al., 2000) and a workload survey (NASA-TLX; Hart & Staveland, 1988). Once these surveys were completed, participants were compensated and debriefed.

Outcome Measures

All participants completed three experimental drives, and their driving behaviors, responses, and other characteristics were recorded from the driving simulator, the eye tracking system, and via survey measures. Participants' scores on the mental model assessment was the key dependent measure for the subsequent modeling.

Drivers' behaviors and other responses were measured using several outcome measures, the selection of which was guided by the measurement matrix (in Part 1) as well as what was practical or feasible in the current study. These outcome measures were grouped into different epochs, based on the experimental tasks and scenarios. The epochs, along with the related measures are described in the sub-sections below.

Normal Driving. In general, normal driving comprised intervals where the participant was driving with or without the ADAS system engaged and did not encounter any scenarios or receive any probes or cues. For each drive, data gathered prior to the first positive acceleration value and after the end of drive marker were excluded. The outcome measures for normal driving are noted in Table 6. Most measures were only aggregated during intervals where the ACC system was active (versus manual driving situations); the exceptions were the measures related to system interactions (system use and system disengagements).

Table 6. Outcome measures for normal driving.

Measure Type	Measure	Definition
Vehicle Control	SD Lane Position	Standard deviation of lane position (in m)
	SD Steering	Standard deviation of angular steer wheel position (in deg)
Safety Event	Crash	Count of crashes with other vehicles or objects
	Lane Departure	Count of instances where lane position exceeded 1.8 meters to the left or right
Attention: Visual Behavior	Mean glance* duration (to forward roadway)	Mean glance duration to forward roadway (in s)
	Glance* count (to forward roadway)	Number of glances to the forward view
	Horizontal and vertical gaze dispersal on forward roadway	Standard deviation of distance of fixation point or eye position/coordinate from center of forward view for glances within this AOI
	Percent dwell time (forward roadway/front screen)	Percentage of time where eyes are directed to the forward roadway relative to all normal epochs
Attention: System Monitoring	Mean glance* durations (to instrument panel [IP])	Mean glance duration to the IP (in s)
	Glance* counts (to IP)	Number of glances to the IP
	Percent dwell time (IP)	Percentage of time where eyes are directed to the IP relative to all other locations within the normal epochs
System Interaction	System disengagements	After initial engagement, number of times the ACC system was disengaged during normal epochs
	System Use	Percentage of time that ACC system was engaged relative to all normal epochs

Note. Unless otherwise noted, measures are aggregated across all normal epochs across the three experimental drives.

**A glance is defined as time from the first fixation to the forward view/IP until the last fixation before the eyes move to another AOI.*

Scenarios. In the time windows surrounding a system transition, driving performance was measured before, during, or after scenarios in which the driver encountered a situation where the driver should resume control from the ACC or where a control transfer was not required (i.e., the ACC was capable of handling these situations). Not all outcome measures were relevant for both Takeover and Non-Takeover scenarios. Table 7 lists those that were investigated for each scenario type.

Table 7. Outcome measures for different scenarios.

Measure	Definition	Scenario Type	
		Non-Takeover[a]	Takeover[b]
Average speed	Mean speed (in mph or m/s)	X	X
Speed variability	Standard deviation of speed during event (in mph or m/s)	X	X
Minimum speed	Minimum speed reached during event (in mph or m/s)	X	X
Maximum speed	Maximum speed reached during event (in mph or m/s)	X	X
Maximum deceleration	Maximum deceleration during event (in m/s/s)	X	X
SD headway	Standard deviation of headway time/distance		X
Minimum headway	Shortest headway time (in s) /distance (in m) during event		X
SDLP	Standard deviation of lane position (in m)	X	X
Maximum lane deviation	Maximum lane position value to the left or right of center (in m)	X	X
SD steering	Standard deviation of steering wheel position (in deg)		X
Maximum steering	Maximum deviation of steering wheel position (in deg)		X
Crash	Crash occurrence (0 = no; 1 = yes)		X
Lane departure	Absolute lane position greater than 1.8 m* during event (0 = no; 1 = yes)		X
Appropriate system transition	Did drivers respond appropriately with respect to system state? (0 = no; 1 = yes; for each scenario type, appropriate response coding was reversed)	X	X
Manner of disengagement	If driver disengaged the system, how did they do so? (1 = brake pedal; 2 = button press; 3 = other)	X	X
Takeover response time	Time from beginning of event until system was disengaged (in s)		X

Measure	Definition	Scenario Type	
		Non-Takeover[a]	Takeover[b]
System accuracy	Given decision to disengage system, did drivers make any interaction errors? (0 = no; 1 = yes)		X
Mean glance durations (to forward roadway)	Mean glance† duration to forward roadway (in s)	X	X
Glance counts (to forward roadway)	Number of glances† to the forward view	X	X
Percent dwell time (forward roadway)	Percentage of time where eyes are directed to the forward roadway relative to all other locations within normal epochs	X	X
Horizontal and vertical gaze dispersal to forward roadway	Standard deviation of distance of fixation point or eye position/coordinate from center of forward view (in m)	X	X
Mean glance durations (to IP)	Mean glance† duration to the IP (in s)	X	X
Glance counts (to IP)	Number of glances† to the IP	X	X
Percent dwell time (IP)	Percentage of time where eyes are directed to the IP relative to all other locations within normal epochs	X	X

[a] Scenarios: Pedestrian (D1S1), Shopping (D1S2), Construction (D2S2), Car on Shoulder (D3S1), Pedestrians on Sidewalk (D3S3), Following Lead Vehicle (D1S5, D2S5, D3S5).

[b] Scenarios: Straddling Vehicle (D1S3), Slow Traffic (D1S4), Oncoming Vehicles (D2S1, D3S2), Fog (D2S4), Oversized Vehicle (D2S3), Bicyclist Group (D3S4).

** This is approximately one-half of the road width.*

†A glance is defined as time from the first fixation to the forward view/IP until the last fixation before the eyes move to another AOI.

Verbal Probes and Cues. Selected outcome measures were gathered during the probe and cue events (see Table 8). These epochs began at the start of the auditory message and lasted between 5 and 34 seconds, depending on the individual probe/cue. The end times were based on a preliminary assessment of the 85th percentile response time from the AAAFTS site.

Table 8. Outcome measures from the probe and cue epochs.

Measure	Definition
Probe/cue accuracy	Correct response to probe or correct execution of cued action (0 = incorrect; 1 = correct)
Mean glance durations (to IP)	Mean glance* duration to the instrument panel (in s)
Glance counts (to IP)	Number of glances* to the IP
Percent dwell time (IP)	Percentage of time where eyes are directed to the IP relative to all normal epochs

** A glance is defined as time from the first fixation to the forward view/IP until the last fixation before the eyes move to another AOI.*

Subjective Measures. Additional measures were gathered through multiple surveys administered throughout the experimental session (see Procedure above) and captured dimensions of driver behavior or perceptions such as trust, understanding, confidence, and acceptance, among others. The following measures were used in the modeling exercise.

- ***Knowledge.*** Mental model (MM) score based on the accuracy of responses to the 20 items in the mental model assessment. This was the response variable used in the modeling.
- ***Confidence.*** Average confidence rating to the 20 items in the mental model assessment.
- ***Trust.*** Average trust rating to the 12 items from the Jian et al. (2000) trust in technology scale. Trust was assessed two times, pre- and post-drive.
- ***Technology Acceptance.*** Average rating across the nine items of the technology acceptance survey (Van der Laan et al., 1997).
- ***Workload.*** Ratings on a seven-point scale gathered from the NASA-TLX for the six sub-scales, including mental demand, physical demand, temporal demand, performance, effort, and frustration (Hart & Staveland, 1988). Each sub-scale was included in the modeling.

Results and Discussion

The initial dataset included 65 participants and 426 variables across all data epochs. Only participants and variables with complete data were retained, and variables with no variation were removed. In order to maximize the available data for modeling, data from both sites was included in the same model. A site variable was included in order to capture any important site-specific variations.

A machine learning approach was used to assess the effectiveness of different outcome variables in predicting driver knowledge of ACC, based on the scores from the mental model assessment. Training (80%) and testing data (20%) sets were extracted from the master data file and a random forest approach was employed. Random forest models have been highly successful in modeling performance and physiological outcomes in other applications related to vehicle automation and driver state (e.g., Han et al., 2023; McDonald et al., 2014).

An overall model was built and evaluated, which included data elements from all data epochs as well as driver demographic features. Separate models were then developed in alignment with the different data epochs:

1. Normal driving
2. Non-takeover scenarios
3. Takeover scenarios
4. Probes and cues
5. Subjective measures

Models for each epoch were run and compared with and without driver demographic and site variables. Although a simulator was used for all of the current data collection, the different models were intended to mimic different data collection capacities. For example, the complexity and challenge of gathering data from surveys, from normal (uneventful) driving, and from situations requiring an overt takeover of control vary considerably. The number of variables or features utilized in the models varied by the epoch, as described in the data section above. Due to technical and coding difficulties, two measures of headway (minimum and variability) were not included in the overall or takeover models.

All model development and evaluation was carried out using R version 4.2-764 (R Core Team, 2024), with the randomForest and caret packages (Liaw & Wiener, 2002; Kuhn, 2008). The optimum number of trees in each model, number of variables tried at each split, and model performance was evaluated using root mean square error (RMSE) and the testing R^2 values.

Model Performance

The model outcomes are shown in Table 9. Several important takeaways are evident. First, the overall model, which included all available data variables across all epochs, performed more poorly (R^2 = 0.18) compared to many of the epoch models that targeted a particular driving situation or type of measure (R^2 = 0.19 to 0.45). This could be due in part to the strength of the trees (i.e., how much model error is reduced by each tree) or high correlation between some of the underlying features. Regardless, this outcome suggests that a more focused array of measures, aligned with different use cases or measurement types, might be more effective predictors of knowledge of vehicle technology in the current study.

Table 9. Model performance for the different driving epochs both with and without driver demographics.

	Without Driver Demographics[a]				With Driver Demographics[a]			
	Variables	Trees/ Split Size[b]	RMSE	R^2	Variables	Trees/ Split Size[b]	RMSE	R^2
Overall	–	–	–	–	395	750/15	2.84	0.18
Normal driving	49	950/30	2.97	0.11	56	1250/56	2.58	0.33
Non-takeover scenarios	143	1450/25	3.02	0.08	150	650/50	2.84	0.19
Takeover scenarios	138	500/50	2.27	0.19	145	500/55	2.34	0.28
Probes/cues	40	950/40	2.95	0.12	47	2000/47	2.34	0.45
Subjective	10	550/5	3.07	0.05	17	500/15	2.35	0.44

[a] *Driver demographics included Age, Gender, Race/Ethnicity, Driving Experience, Weekly Miles, Previous ACC Experience (yes/no), and Site.*

[b] *The number of trees and variables tried at each split was based on the highest R^2 and lowest RMSE for each, checking at intervals of 50 for trees sized 500–2000, and checking the number of variables tried at each split from 5–75 in intervals of 5.*

Second, the models performed much better when informed by *some* information about the drivers; that is, including certain demographic information enhanced model performance significantly. For example, epoch model performance ranged from R^2 = 0.05 to 0.19 when driver information was withheld. When included, model performance ranged from R^2 = 0.19 to 0.45. Thus, it is prudent to consider tailored or personalized models whenever possible (e.g., Han et al., 2023). Even so, the performance of the current models suggests that there are other unmeasured factors that could account for additional variance in the models.

Third, absent demographic and driver-related variables, the takeover scenario epoch model performed best (by a small margin), which included variables related to how drivers interact with the technologies in and around situations that the system is nearing its limitations. When demographic features were included, however, the subjective and probes/cues epoch models performed best, followed by takeover scenarios and normal driving epochs (see Table 9). Even with the inclusion of demographic information, the non-takeover scenario epoch model continued to perform poorly, suggesting that mundane or non-events—at least those contrived in the current context—might not offer significant insight into driver knowledge of vehicle systems (though they are still an integral part of a robust experiment or evaluation, e.g., to provide space and uncertainty around critical events and the like).

Features of Importance

Beyond the model performance for the different epochs and data types, the most relevant features of importance were also examined in order to explore which underlying data elements might be most useful in predicting driver knowledge of vehicle technology. In each model, variable importance scores were determined by each variable's mean decrease in accuracy (as defined by Liaw & Wiener, 2022), or how much the accuracy of the model decreases without that variable. In Figures 4–8 below, the top model features are plotted for each data epoch, with side-by-side comparisons of the models with and without demographic variables.

For the Normal driving epoch, the most important features for both models largely related to the allocation of visual attention to the roadway and to the instrument cluster (see Figure 4), and more specifically, to the number of glances, the glance durations, and the resulting attention ratio (distribution of glances to the roadway and instrument panel). Standard deviation of steering was the only vehicle control–related metric that figured in the top ten features.

For the Non-Takeover Scenario epochs, measures related to visual attention were most prominent (Figure 5). Maximum lane departure ranked highly, possibly indicating the adoption of greater safety margins in non-transition cases (e.g., moving further to the left side of lane when approaching pedestrians on the right side). In terms of the scenarios, following a lead vehicle, passing a pedestrian group, and passing a construction zone yielded the most important model features when driver demographics were not included in the model. When these features were included, the parked car and shopping area scenarios also contributed some of the most important features.

With respect to Takeover epochs, while glance metrics continued to be prominent model features, more vehicle control and safety outcomes were represented (Figure 6). For example, measures of velocity and deceleration as well as the occurrences of crashes were represented in these takeover situations. These features originated from scenarios involving slow or erratic lead vehicles, a group of bicyclists, oncoming vehicles, or from the encounter of fog.

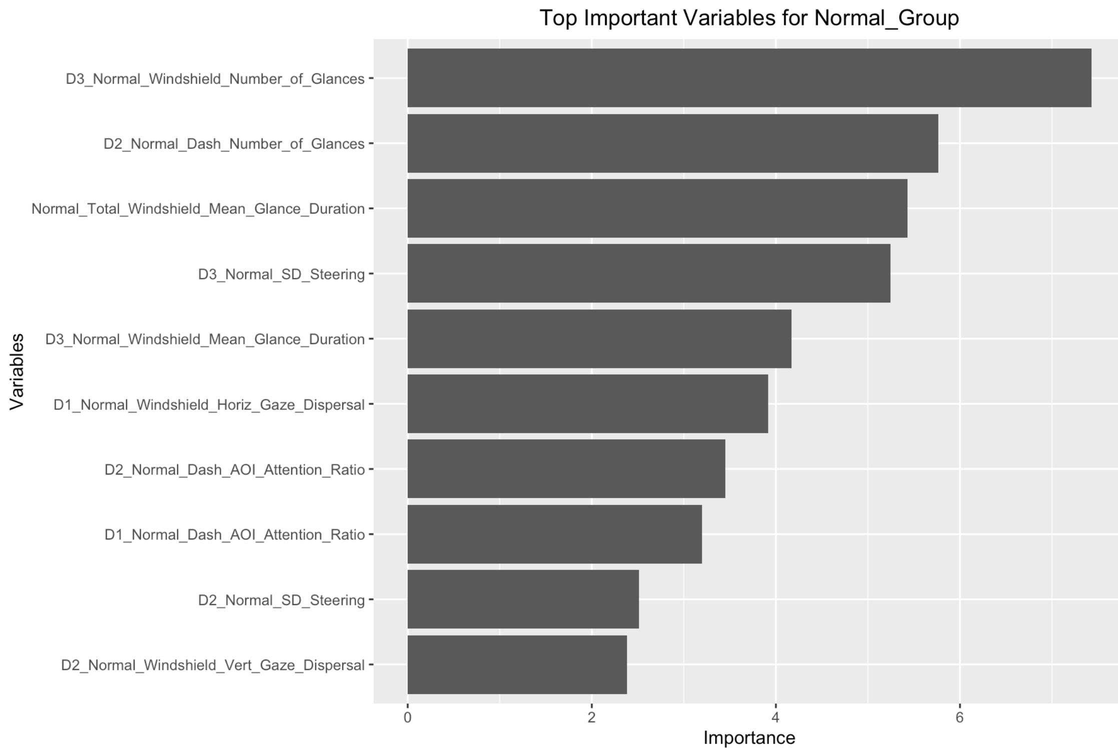
For Probe and Cue epochs, the most important features were almost exclusively those related to how drivers distributed their attention to the roadway and instrument panel (Figure 7). Performance (accuracy) variables also featured in the model without driver demographics.

With respect to Subjective measures, confidence in knowledge, frustration, technology acceptance, and trust were the most prominent model features (Figure 8). Interestingly, confidence was the top performing feature in the current study, although other recent studies have found that confidence often is dissociated with knowledge of vehicle technology (e.g., Lenneman et al., 2020; Mason et al., 2023; Carney et al., 2022).

Overall, measures of visual attention were the most important features across all of the epochs examined (excluding Subjective). This suggests that glance behavior has promise as a proxy for driver knowledge of vehicle technology. Indeed, how drivers distribute their attention to the roadway and instrument panel logically should reflect their engagement in the task and their knowledge of their underlying responsibilities to continue to monitor the traffic environment as well as the system status.

With few exceptions, vehicle control measures did not feature prominently in the modeling efforts. While this could reflect the relatively straight roadways and traffic environments adopted in the current study, it could also underscore the role and influence of other factors unrelated to knowledge of technology in determining momentary vehicle control. That said, vehicle control measures (and, similarly, safety outcomes) did feature more prominently in the takeover scenarios, where an overt response was required in order to successfully avoid a traffic conflict. In general, driving experience, race/ethnicity, and, to a lesser extent, age were the most important demographic features when those variables were included.

Without Driver Demographics



With Driver Demographics

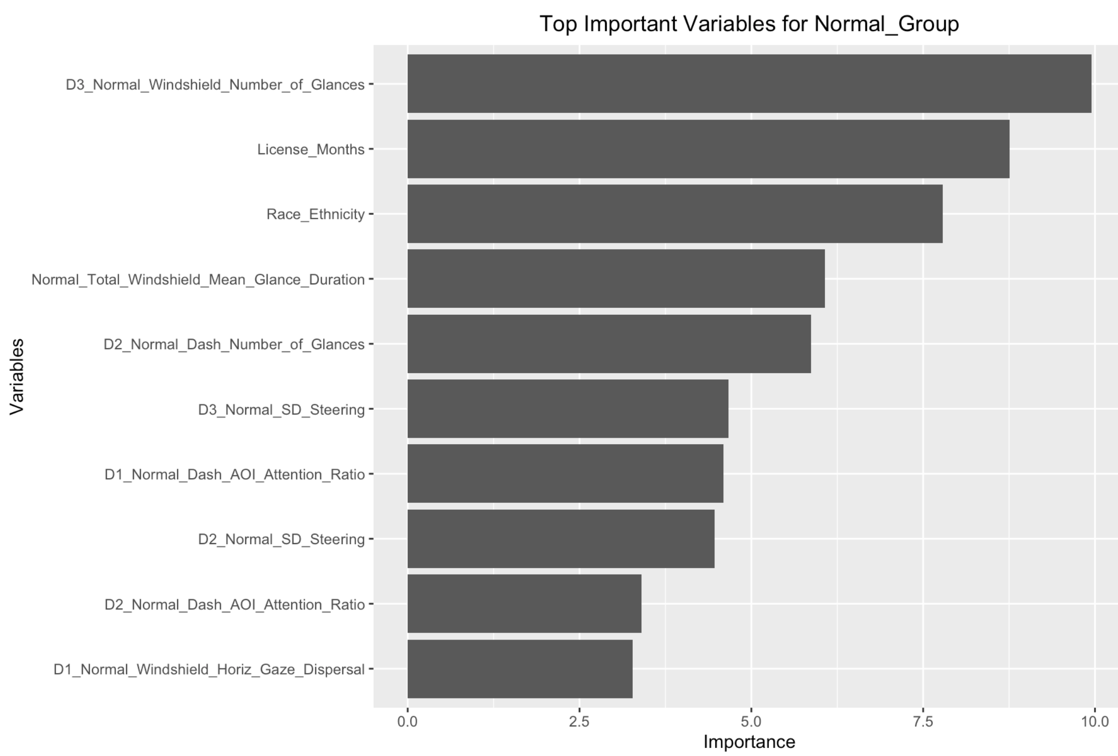
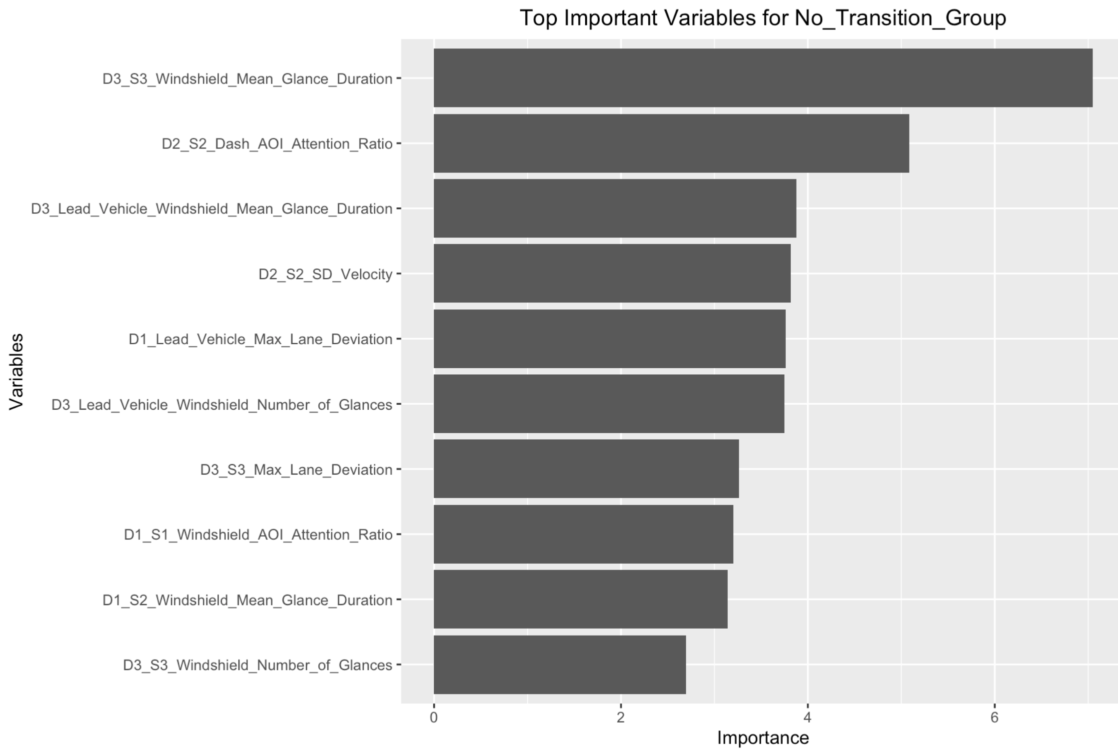


Figure 4. Top features of importance for Normal Driving epoch.

Without Driver Demographics



With Driver Demographics

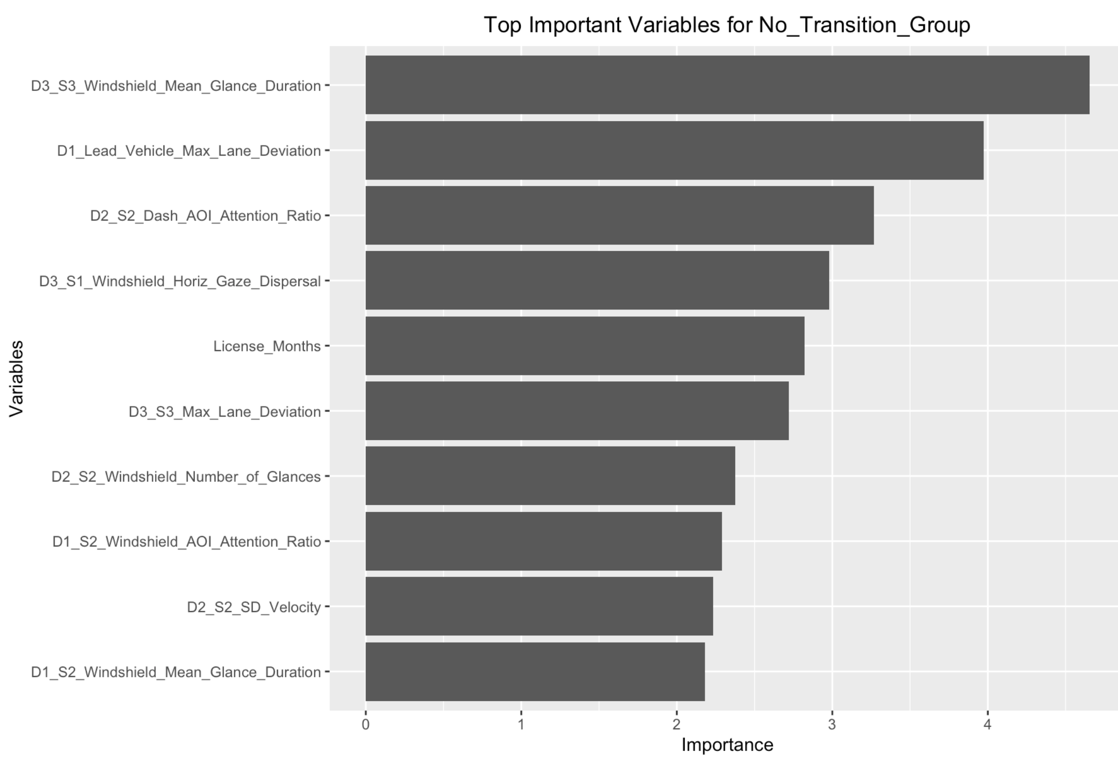
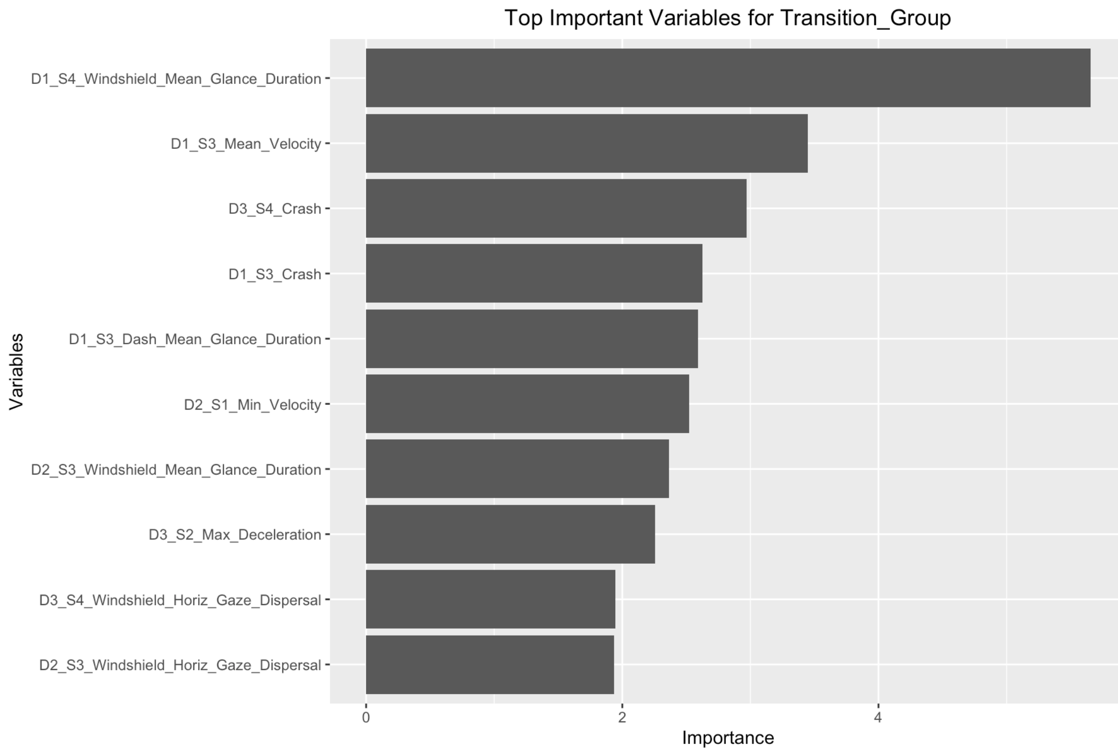


Figure 5. Top features of importance for Non-Takeover Scenario epoch.

Without Driver Demographics



With Driver Demographics

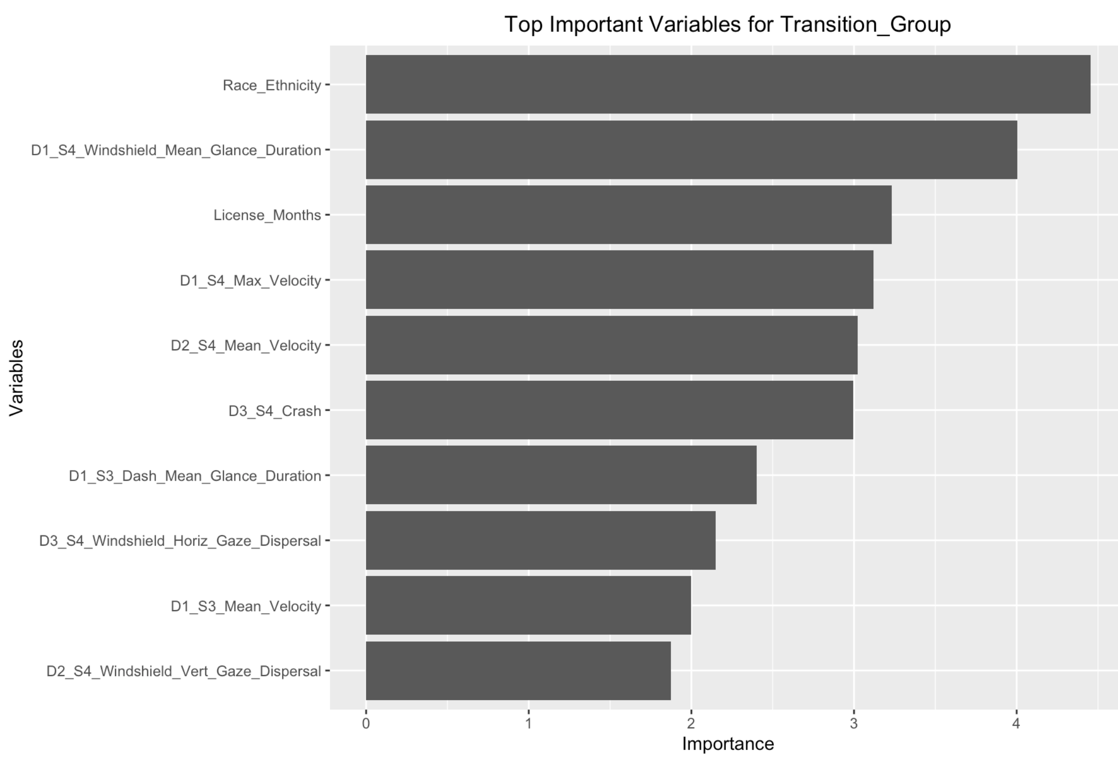
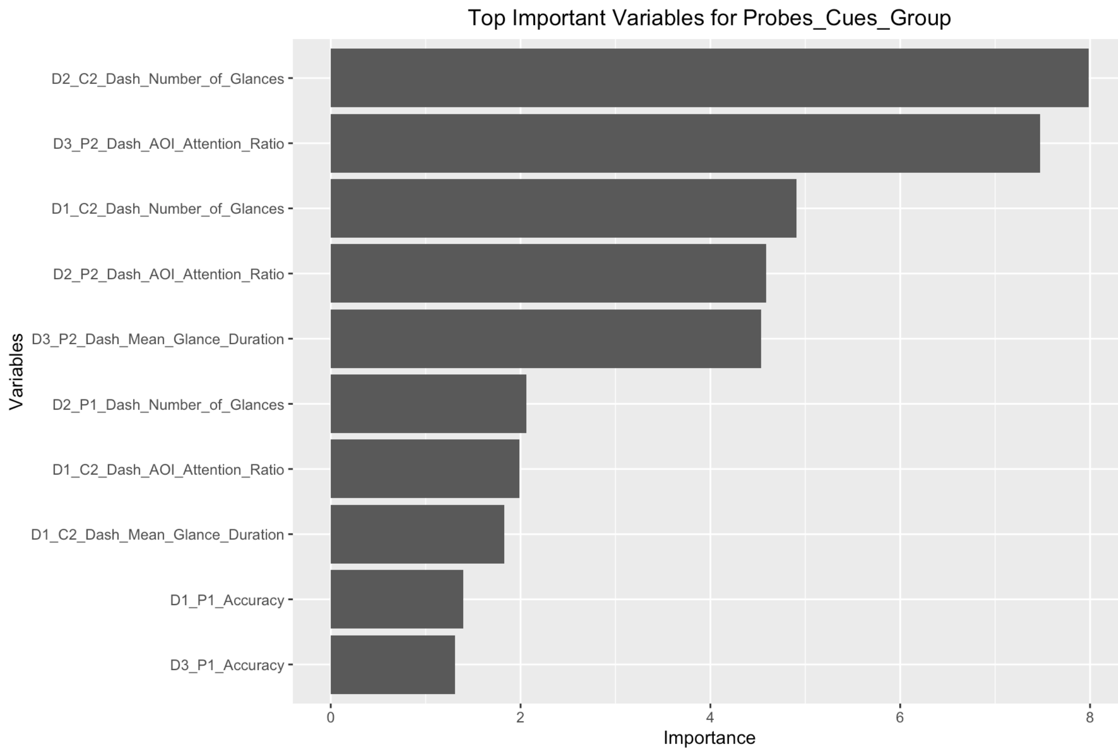


Figure 6. Top features of importance for Takeover Scenario epochs.

Without Driver Demographics



With Driver Demographics

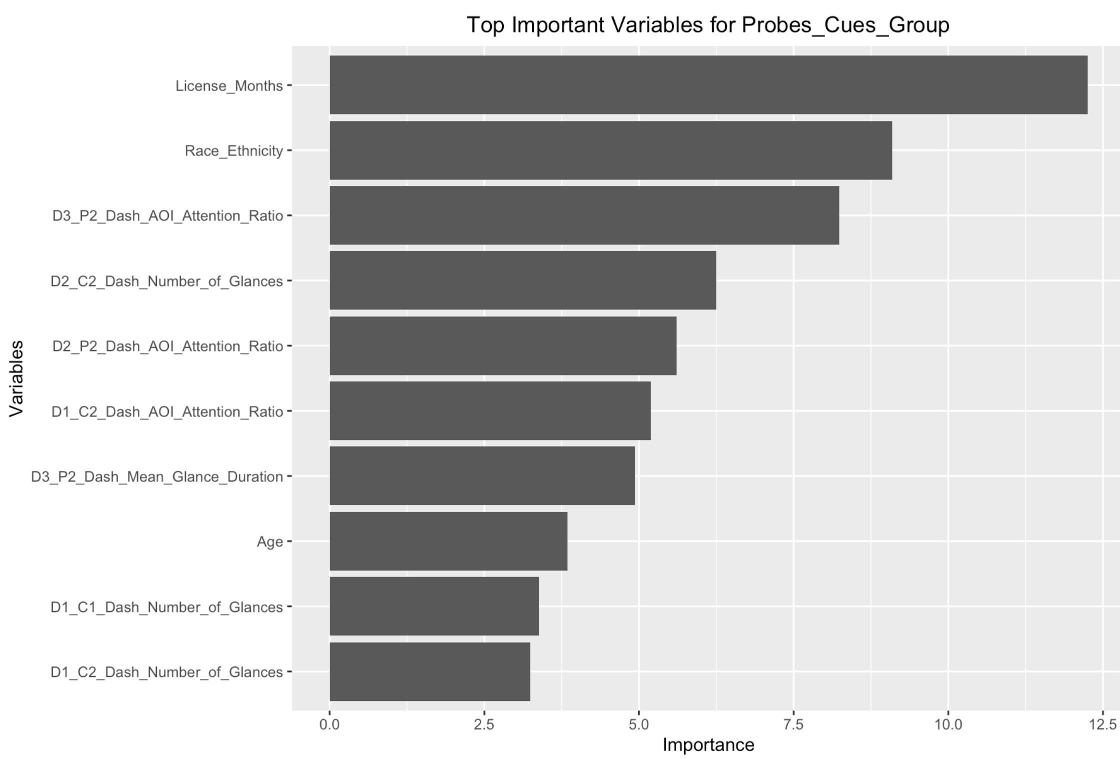
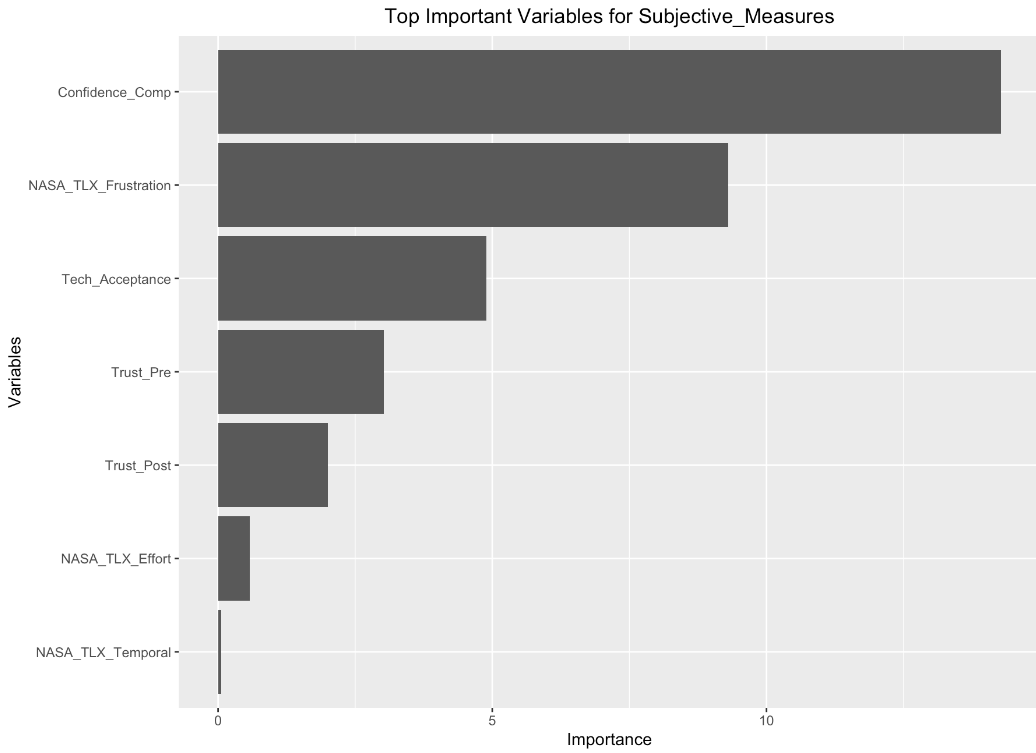


Figure 7. Top features of importance for Probe / Cue epochs.

Without Driver Demographics



With Driver Demographics

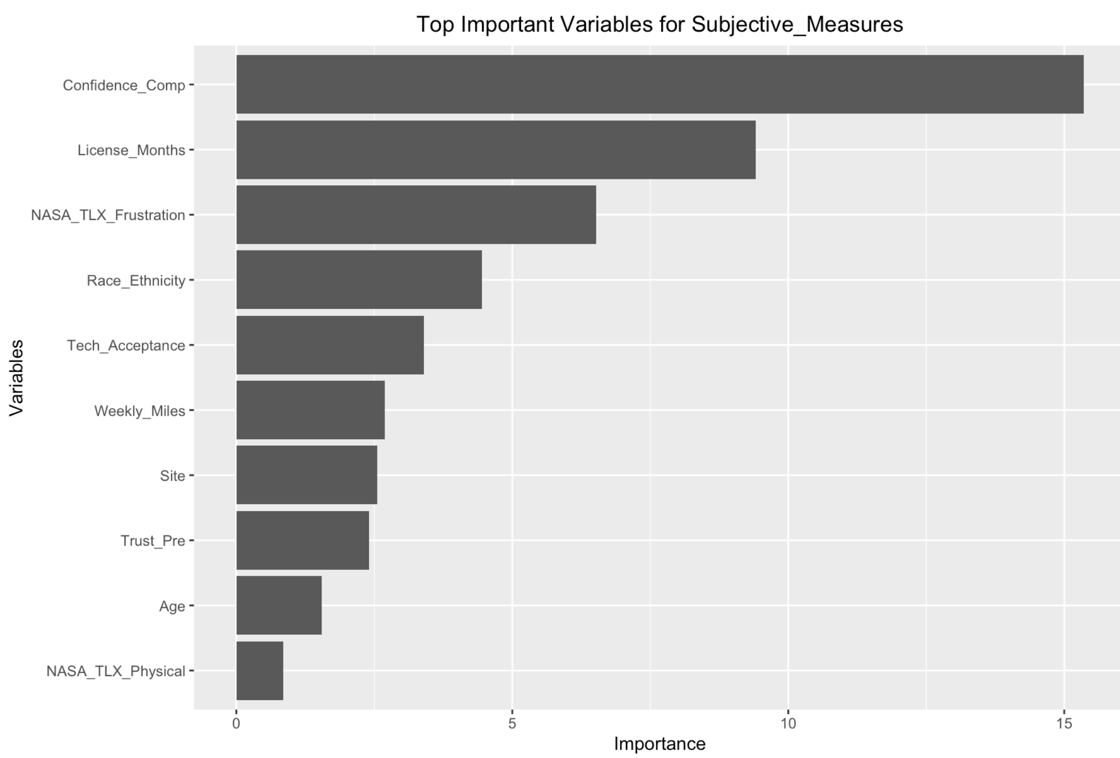


Figure 8. Top features of importance for Subjective measures epoch.

General Discussion and Key Takeaways

Advanced vehicle technologies have the potential to improve safety and convenience for drivers (e.g., Bengler et al, 2014; Fisher et al., 2020); however, in order to realize these benefits, it is important that drivers use the systems appropriately. Past research has documented gaps in drivers' and other road users' understanding of how new technology works and when it should be used (e.g., McDonald et al., 2018; Gaspar et al., 2020; Lenneman et al., 2020; Mason et al., 2023). Thus, the interplay between driver knowledge, performance and safety outcomes, and approaches for training and education have become topics of interest in the research and stakeholder communities. Along these lines, further work exploring the relationship between driver knowledge and outcomes on a wide variety of performance and safety measures is needed. This would offer more tools to establish or infer a driver's understanding of technology as well as instruments to assess the efficacy of different approaches to training and education.

Thus, the overall objective of this study was to identify, catalog, and appraise outcome measures in terms of their utility in assessing or inferring driver knowledge of advanced vehicle technologies. The outcomes of the two parts of the study, the review of potential measures and the experimental validation study, are summarized briefly below followed by key takeaways and caveats.

Landscape of Outcome Measures

The purpose of this review was to enumerate and catalog measures of performance, safety, and behavior in the driving domain that could potentially bear on driver knowledge of advanced vehicle technology. The review of the scientific literature, engineering standards, and existing datasets yielded a wide array of measures, which were distilled into a smaller set of nine categories. These categories are as follows:

- Control transitions
- System interactions
- Vehicle control
- Safety events
- System understanding
- Attitudes and perceptions
- Workload, Usability
- Attention

A detailed accounting of each category is provided in the matrix above and in an alternative format here.

Overall, the measures varied greatly in terms of how they might be implemented and how challenging the implementation would be. Each category had noteworthy advantages and disadvantages with respect to their bearing on system knowledge and, in many cases, the relationship between the measure and driver knowledge had not been investigated empirically.

Direct measurement or assessments of driver knowledge through surveys or other instruments were considered a gold standard; however, with some shortcomings related to the feasibility of measuring over time. Many measures related to control transitions and system interactions were thought to be beneficial as they reflect appropriate and safe use of the systems. These have been examined empirically in some past research as well (e.g., Gaspar et al., 2020). In contrast, vehicle control and safety measures, such as speed, lane keeping, and crashes, are commonly used in safety research, but are impacted by other factors besides driver knowledge, including environmental or situational factors. In addition, crashes and near misses, including lane departures, are very rare events, which detract from their usefulness in appraising driver knowledge of vehicle technology or learnings from training and educational approaches.

Measures of attitudes and perceptions can be easily administered and, while they often deal with constructs that are related to knowledge of technology, they are also impacted by other factors. Similarly, measures of workload can shed insight into the driver's state in different situations but are often susceptible to noise and other extraneous sources.

Driver attention is often assessed in research settings and can also be implemented in situ. While the distribution of one's attention is shaped by other factors (as for other measures noted above), one discernible benefit is that driver attention can be mapped directly to critical knowledge regarding driver responsibilities while using certain vehicle technology. That is, eye glance metrics can be good indicators that a driver is appropriately engaged in traffic scanning behavior and system monitoring when the system is engaged. Moreover, visual scanning is a skill that can be trained (e.g., Pradhan et al., 2005; 2011; Unverricht et al., 2018).

All in all, the outcomes from Part 1 suggest that combinations of outcome measures could offer a better indication of a driver's underlying knowledge, owing to the many pros and cons articulated in the matrix. That is, adopting multiple measures would help offset some of the inherent shortcomings of just a single measure or class of measures. The next phase of the project sought to better elucidate the relationship between a subset of these measures and system knowledge.

Relationship among Outcomes and System Knowledge

The purpose of Part 2 was to identify and validate outcome measures that can be implemented in research to measure drivers' knowledge and understanding of advanced vehicle systems using a multi-site experimental approach. A driving simulator study was executed at two different sites using a common approach, scenarios, and measures. While a full accounting of all measures identified in Part 1 was infeasible, the study sought to include as many as possible as well as incorporating different use cases (i.e., types of measurement epochs or test windows).

The results from the random forest modeling suggested that adopting a more stringent set of measures (according to measurement epochs and types) was preferred to including every available measure—at least was in the case in the current context where the number of available measures likely exceeded what might be present in other settings.

Importantly, having some demographic information about the driver helped to enhance the performance of the models and, by extension, the utility of the outcome variables in predicting system knowledge. In many though perhaps not all use cases, including some information about the driver appears feasible.

Depending on the availability of demographic and driver-related variables, measures that involve routine interactions with the systems (e.g., probes and cues) or in control transition or takeover situations contributed to the model performance, as did other complimentary subjective measures and measures gleaned during normal system operation (i.e., normal driving epochs).

With respect to more specific metrics, measures of visual attention were generally the most important features in the modeling efforts. That is, the manner in which drivers distribute their attention to the roadway and instrument panel appears to be indicative of their knowledge of their underlying responsibilities or of the system limitations. In contrast, most measures of vehicle control failed to provide any important insight in the predictive models. This could reflect the influence of other factors unrelated to driver knowledge of technology in determining momentary vehicle control. That said, vehicle control and safety measures were relevant when examining control transition situations, where active control was required to avoid conflict.

Takeaways

Collectively, the two parts comprising this research study shed some insight into the utility of different types of measures in assessing driver knowledge of advanced vehicle technology. Ultimately, employing a cluster of measures is advised in this space in order to account for some of the limitations associated with an individual measure.

While the matrix offered some high-level perspectives on when a certain measure might be assessed (e.g., pre-drive or in situ), the results from the experimental study offered more specific information about the optimal timing of measurements in relation to different situations or use cases. That is, the experimental work further highlights how the variables most strongly related to knowledge of ADAS vary by measurement window; in particular, those measurement windows or epochs involving system interactions or control takeovers, or routine situations where the system is operating showed stronger outcomes in the modeling exercise. This is useful information for those who might be constrained by the type of data that can be acquired or assessment that can be conducted.

Eye glance measures also featured prominently in all of the relevant models, including those related to glance location, frequency, and duration. This outcome corroborated some of the perspectives gleaned from the matrix noting the advantages of eye glance metrics in mapping onto a driver's roles and responsibilities while using the technology. More work is needed to establish whether head movements, which are easier to measure in situ, might serve as a sufficient proxy for eye glance measures.

With respect to measures of vehicle control, the matrix highlighted some constraints in their use as these measures are often impacted by many other factors. This too was evident from the experimental study, with vehicle control and related safety measures not showing as prominent components of the models in most cases. The noteworthy exception was in cases where active vehicle control was necessary (i.e., control transition or takeover scenarios).

Subjective measures, such as confidence and acceptance of technology and trust, were also good predictors of knowledge. However, practically speaking, if one were in a position to gather an array of subjective measures, it is plausible that measuring knowledge of vehicle technology directly would also be practicable. What begs further research is whether some subjective measures could be distilled in a shorter form, equally useful or predictive, that could be captured more effectively in situ. For example, whether a survey instrument with several or many items could be equally effective with one or two key items.

In summary, researchers and other stakeholders interested in gauging driver knowledge of technology or in assessing educational or training approaches should do the following:

- Prioritize measures of eye glance behavior to corroborate a driver's knowledge of the vehicle system; vehicle control and safety measures are more effectively applied in specific edge-case or takeover situations.
- Establish measurement windows around system interactions (e.g., changing system settings), takeover situations, and normal system operation (where the system is activated).
- Leverage subjective measures, such as confidence or technology acceptance whenever possible, which are strong predictors of knowledge.
- Incorporate or consider information about the driver, such as driving experience and demographics.

It is also important to underscore that tying performance outcomes to driver knowledge of vehicle technology and, by extension, to the efficacy of training and educational efforts is but one lens through which to interpret the current outcomes. A broad assessment of behavior and performance, irrespective of knowledge and reflecting the multifaceted dimensions of the matrix, can be useful in its own right as this can lend insight into the overall safety of an individual or groups of drivers.

Limitations and Future Directions

While the current study offers some useful insight into proxies for driver knowledge of vehicle technologies, there are a number of limitations that should be noted. First, the experimental study was in a simulator setting, with limited range of variables and a limited array of driving conditions. It was not feasible to test the entire universe of possible outcome measures, so the approach was aimed at gathering a useful representation of measures or categories of the matrix. However, it is not exhaustive and there are questions about how well the current results can generalize to other settings or use cases. Moreover, some categories of measures, such as system interaction, were represented by only a handful of measures whereas other categories had many measures (e.g., attention and vehicle control). While this is a natural by-product of the types of measures themselves, future work could explore a broader array of measures according to some of the matrix categories. This could also help inform on other relevant factors that could increase the performance of the models (e.g., increasing the R^2 values). Future work can also consider how different measures evolve over time as drivers gain more experience with the systems.

The current study focused on one ADAS technology, adaptive cruise control, as this technology is widely available yet often misunderstood (e.g., McDonald et al., 2018). Moreover, the current study only implemented a single version of ACC; it is very likely that the design of the system could impact many of the specific measures and system knowledge itself (e.g., more intuitive designs). It is unclear whether the current outcomes share implications for other forms of technology and higher levels of automation. While this is an open question, it is believed that much of the broad foundational work expressed in the matrix can be applied to other forms of ADAS technology, as well as to more advanced levels of automation. It should be noted too that, on the other end of the spectrum, understanding of active collision avoidance systems, such as automatic emergency braking (AEB), is less important given that such systems are intended to intervene only in critical situations and do so without any driver input. The nuance of individual variables accounted for in the experimental work might vary for other forms of technology, however, and should be the focus of further investigation.

Another interesting question is whether different measures (or the ideal combinations) could be tailored to assess different components of understanding. The current study used a single score to denote overall knowledge of the system; however, this could be distilled into subcomponents such as knowledge related to operation of the system versus knowledge of system capabilities or limitations. For example, measures of system interactions might map well onto understanding of system function and operation, whereas visual attention might be more reflective of a driver's understanding of system limitations.

Lastly, while the current work is often related or extrapolated to approaches to training and education, more work is needed to empirically examine the direct impact of different training or educational materials on different outcome measures, the selection of which could be informed by the current work. This could also complement a broader discussion and consideration of what are the overall aims of any training or educational approach.

References

Bengler, K., Dietmayer, K., Farber, B., Maurer, M., Stiller, C., & Winner, H. (2014). Three decades of driver assistance systems: Review and future perspectives. *IEEE Intelligent Transportation Systems Magazine* 6(4), 6–22. doi: 10.1109/MITS.2014.2336271

Carney, C., Gaspar, J. G., Roe, C. & Horrey, W.J. (2022). *An Examination of How Longer-Term Exposure and User Experiences Affect Drivers' Mental Models of ADAS Technology* (Technical Report). Washington, D.C.: AAA Foundation for Traffic Safety.

DeGuzman, C. A., & Donmez, B. (2022). Drivers don't need to learn all ADAS limitations: A comparison of limitation-focused and responsibility-focused training approaches. *Accident Analysis & Prevention* 178, 106871. doi: 10.1016/j.aap.2022.106871

Fisher, D. L., Horrey, W. J., Lee, J. D., & Regan, M. A. (2020). *Handbook of Human Factors for Automated, Connected, and Intelligent Vehicles*. Boca Raton, FL: CRC press.

Forster, Y., Hergeth, S., Naujoks, F., Krems, J., & Keinath, A. (2019). User education in automated driving: Owner's manual and interactive tutorial support mental model formation and human-automation interaction. *Information* 10(4), 143. doi: 10.3390/info10040143

Forster, Y., Hergeth, S., Naujoks, F., Krems, J. F., & Keinath, A. (2020). What and how to tell beforehand: The effect of user education on understanding, interaction and satisfaction with driving automation. *Transportation Research Part F: Traffic Psychology and Behaviour* 68, 316–335. doi: 10.1016/j.trf.2019.11.017

Gaspar, J. G., Carney, C., Shull, E. & Horrey, W.J. (2020). *The Impact of Driver's Mental Models of Advanced Vehicle Technologies on Safety and Performance* (Technical Report). Washington, D.C.: AAA Foundation for Traffic Safety.

Han, D., Liu, J., Zhou, F., Tilbury, D., Robert, L., Molnar, L. & Yang, X. J. (2023). *Measuring and Predicting Drivers' Takeover Readiness and Supporting Takeover Transitions in Automated Driving* (Technical Report). Washington, D.C.: AAA Foundation for Traffic Safety.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology* 52, 139–183. doi: 10.1016/S0166-4115(08)62386-9

Jian, J. Y., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics* 4(1), 53–71. doi:10.1207/S15327566IJCE0401_04

Kuhn, Max (2008). "Building Predictive Models in R Using the caret Package." *Journal of Statistical Software 28*(5), 1–26. doi:10.18637/jss.v028.i05

Lenneman, J., Mangus, L., Jenness, J., & Petraglia, E. (2020). Delineating clusters of learners for driver assistance technologies. *HCI International 2020 – Late Breaking Posters. HCII 2020. Communications in Computer and Information Science*, 1294. Springer, Cham. doi: 10.1007/978-3-030- 60703-6_77

Liaw A, & Wiener M (2002). "Classification and Regression by randomForest." *R News* 2(3), 18–22. https://CRAN.R-project.org/doc/Rnews/.

Mason, J., Carney, C., Gaspar, J. G., Kim, W., Romo, A. & Horrey, W.J. (2023). *Mapping Comprehension of ADAS across Different Road Users* (Technical Report). Washington, D.C.: AAA Foundation for Traffic Safety.

McDonald, A., Carney, C. & McGehee, D.V. (2018). *Vehicle Owners' Experiences with and Reactions to Advanced Driver Assistance Systems* (Technical Report). Washington, D.C.: AAA Foundation for Traffic Safety.

McDonald, A. D., Lee, J. D., Schwarz, C., & Brown, T. L. (2014). Steering in a random forest: Ensemble learning for detecting drowsiness-related lane departures. *Human Factors* 56(5), 986–998. doi: 10.1177/0018720813515272

Naumann, R.B., Kreuger, L. K., Sandt, L., Hassmiller Lich, K., Bauchwitz, B., Kumfer, W. & Combs, T. (2023). *Examining the Safety Benefits of Partial Vehicle Automation Technologies in an Uncertain Future* (Technical Report). Washington, D.C.: AAA Foundation for Traffic Safety.

Pradhan, A. K, Divekar, G., Masserang, K., Romoser, M., Zafian, T., Blomberg, R. D., Thomas, F. D., Reagan, I., Knodler, M., & Pollatsek, A. (2011). The effects of focused attention training on the duration of novice drivers' glances inside the vehicle. *Ergonomics* 54(10), 917–931. doi: 10.1080/00140139.2011.607245

Pradhan, A. K., Fisher, D. L., & Pollatsek, A. (2005). The effects of PC-based training on novice drivers' risk awareness in a driving simulator. *Proceedings of the Third International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design* 3, 81–87. doi: 10.17077/drivingassessment.1146

R Core Team (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.Rproject.org/

Rasmussen, J. (1983). Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man, and Cybernetics* SMC-13(3), 257–266. doi: 10.1109/TSMC.1983.6313160

Singer, J. & Jenness, J.W. (2020). *Impact of Information on Consumer Understanding of a Partially Automated Driving System* (Technical Report). Washington, D.C.: AAA Foundation for Traffic Safety.

Unverricht, J., Samuel, S., & Yamani, Y. (2018). Latent hazard anticipation in young drivers: Review and meta-analysis of training studies. *Transportation Research Record* 2672(33), 11–19. doi: 10.1177/0361198118768530

Van der Laan, J.D., Heino, A., & De Waard, D. (1997). A simple procedure for the assessment of acceptance of advanced transport telematics. *Transportation Research Part C: Emerging Technologies* 5(1), 1–10. doi: 10.1016/S0968-090X(96)00025-3

Yang, C.Y.D., Horrey, W.J., Tefft, B.C. & Kim, W. (2023). *User Interactions with Vehicle Automation Technologies: A Review of Previous Research and a Proposed Framework* (Research Brief). Washington, D.C.: AAA Foundation for Traffic Safety.

Appendix A: Training Materials (after Gaspar et al., 2020)

Novice Group (UMass Site)

Adaptive Cruise Control (ACC)

- Assists the driver with maintaining a set speed and a set gap from the vehicle in front of them.
- If no vehicle is detected, the system works like conventional cruise control and maintains a set speed. If you, the driver, encounter a vehicle travelling slower, the ACC will reduce your vehicle's speed and maintain the fixed gap between your vehicle and the vehicle ahead. When the slower moving vehicle is no longer in front of you, the system will accelerate your vehicle in order to resume the set speed.

Steps to Activate ACC

1. Press the "ON" button on the steering wheel to turn ACC 'On'.
2. To adjust speed, press the "SET +" or "SET -" button on the steering wheel.
3. To adjust following gap, press the "OK" button on the steering wheel.

Steps to Cancel & Resume ACC

- To turn ACC 'Off' (no longer active): press the "OFF" button or depress the brake pedal
- To reactivate from the 'Off' state, you will need to press the "ON" button, then set speed and following gap.
- To place ACC in 'Standby' rather than 'Off': press the "CNCL" button or depress the gas pedal
- To resume ACC from the 'Standby' state, press the "RES" or "SET +" button on the steering wheel

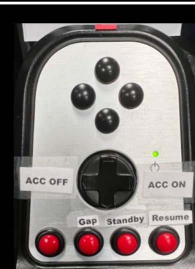
Novice Group (AAAFTS Site)

Adaptive Cruise Control (ACC)

- Assists the driver with maintaining a set speed and a set gap from the vehicle in front of them.
- If no vehicle is detected, the system works like conventional cruise control and maintains a set speed. If you, the driver, encounter a vehicle travelling slower, the ACC will reduce your vehicle's speed and maintain the fixed gap between your vehicle and the vehicle ahead. When the slower moving vehicle is no longer in front of you, the system will accelerate your vehicle in order to resume the set speed.

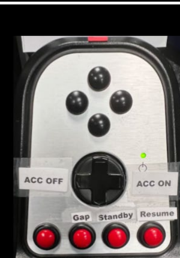
Steps to Activate ACC

1. Press the Right D-pad button to turn ACC 'On'.
2. Press the Up or Down D-pad button to set the ACC speed (ACC activated).
3. To adjust speed, press the Up or Down D-pad button in the direction needed.
4. To adjust following gap, press the second red button on the console.



Steps to Cancel & Resume ACC

- To turn ACC 'Off' (no longer active): Press the Left D-pad button
- To reactivate from the 'Off' state, you will need to press the Right D-pad button, then set speed and following gap.
- To place ACC in 'Standby' rather than 'Off': 2 methods
  - Press the brake
  - Press the 3rd red button on the console
- To resume ACC from the 'Standby' state, press the fourth red button on the console



Advanced Group (UMass Site)

Adaptive Cruise Control (ACC)

- Assists the driver with maintaining a set speed and a set gap from the vehicle in front of them.
- If no vehicle is detected, the system works like conventional cruise control and maintains a set speed. If you, the driver, encounter a vehicle travelling slower, the ACC will reduce your vehicle's speed and maintain the fixed gap between your vehicle and the vehicle ahead. When the slower moving vehicle is no longer in front of you, the system will accelerate your vehicle in order to resume the set speed.

Adaptive Cruise Control (ACC)

- The driver selects the following gap and can adjust it whenever ACC is activated.
- ACC can be activated at speeds of 30 mph and above.
- If the accelerator is pressed when ACC is activated (set speed), it will not be able to brake to keep you within the desired set following gap.
- ACC is primarily intended for driving on dry, straight roads, such as highways and freeways. It should not be used on city streets or when traffic is constantly changing.

Adaptive Cruise Control (ACC)

- ACC typically uses radar to detect vehicles in front of you and to estimate the following gap.
- ACC may not detect a vehicle even when a vehicle is present in the lane ahead:
  - Roadway elevation changes
  - Turning vehicles
  - Smaller vehicles like motorcycles
  - Vehicles not in the center of the lane ahead

Advanced Group (AAAFTS Site)

Adaptive Cruise Control (ACC)

- Assists the driver with maintaining a set speed and a set gap from the vehicle in front of them.
- If no vehicle is detected, the system works like conventional cruise control and maintains a set speed. If you, the driver, encounter a vehicle travelling slower, the ACC will reduce your vehicle's speed and maintain the fixed gap between your vehicle and the vehicle ahead. When the slower moving vehicle is no longer in front of you, the system will accelerate your vehicle in order to resume the set speed.

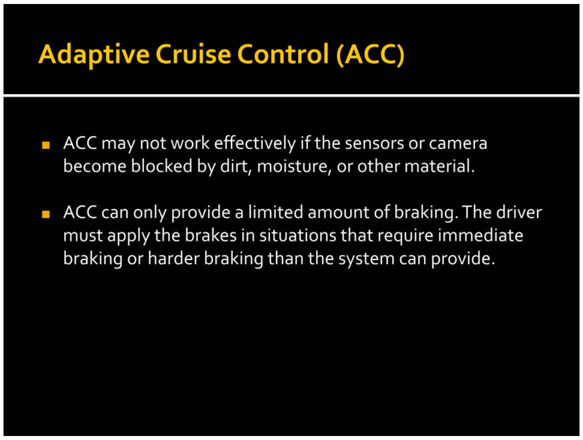
Adaptive Cruise Control (ACC)

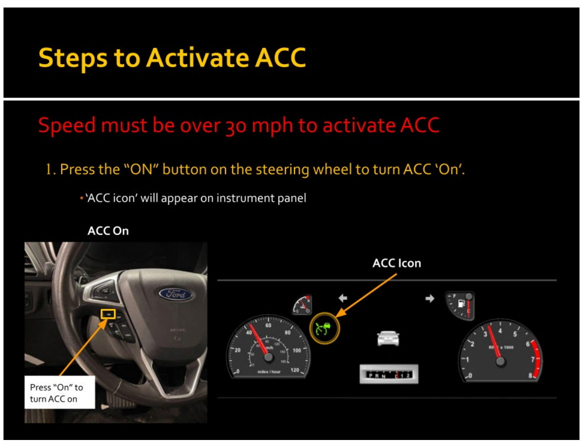
- The driver selects the following gap and can adjust it whenever ACC is activated.
- ACC can be activated at speeds of 40 mph and above.
- If the accelerator is pressed when ACC is activated (set speed), it will not be able to brake to keep you within the desired set following gap.
- ACC is primarily intended for driving on dry, straight roads, such as highways and freeways. It should not be used on city streets or when traffic is constantly changing.

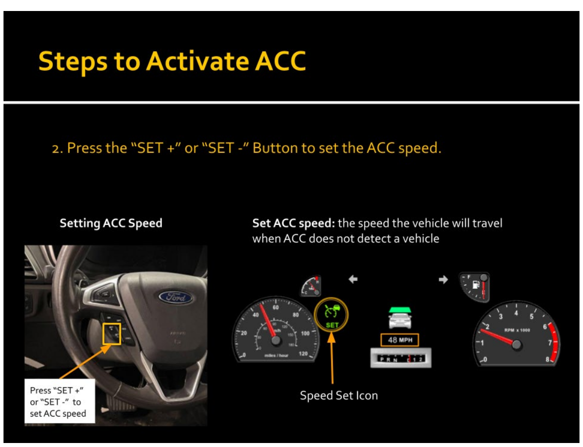
Adaptive Cruise Control (ACC)

- ACC typically uses radar to detect vehicles in front of you and to estimate the following gap.
- ACC may not detect a vehicle even when a vehicle is present in the lane ahead:
  - Roadway elevation changes
  - Turning vehicles
  - Smaller vehicles like motorcycles
  - Vehicles not in the center of the lane ahead

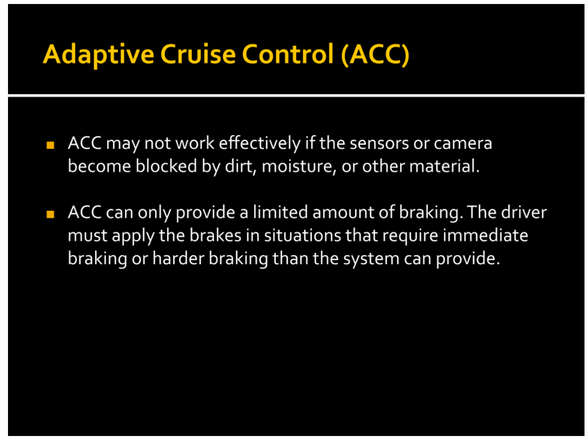
Advanced Group (UMass Site)

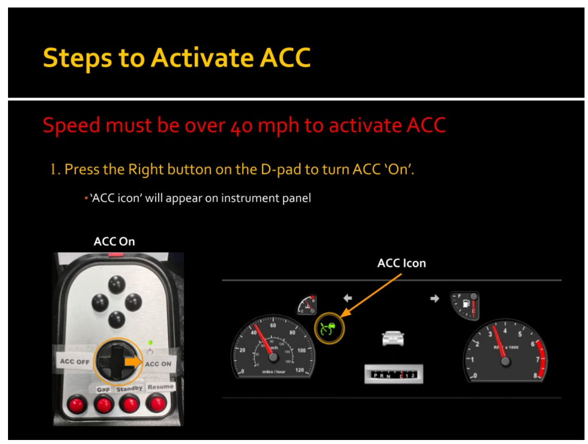




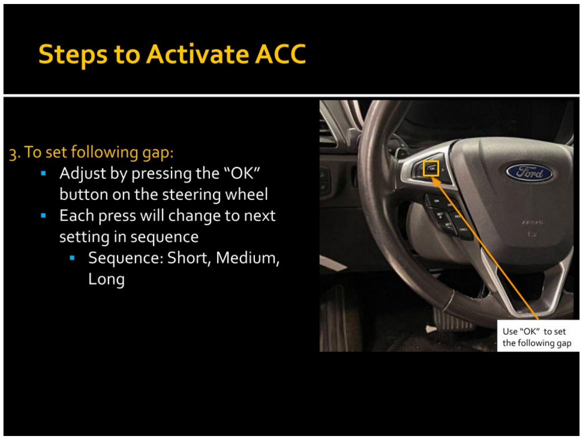


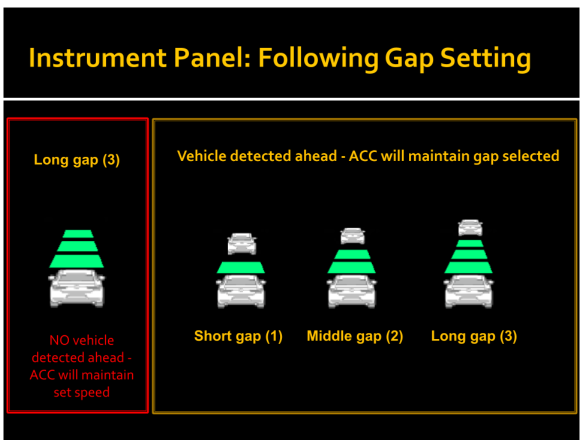
Advanced Group (AAAFTS Site)

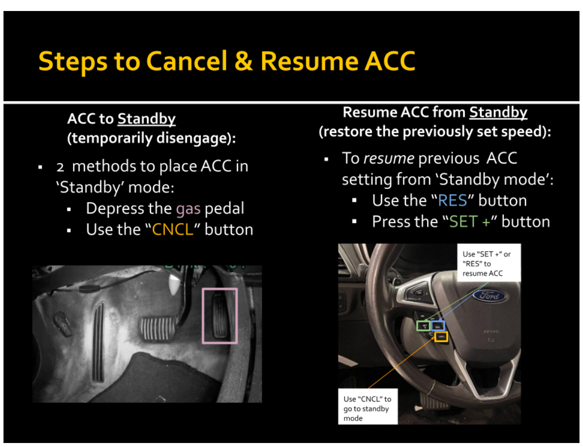




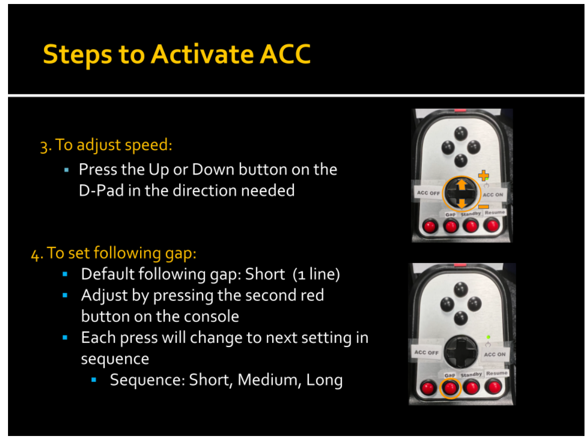
Advanced Group (UMass Site)

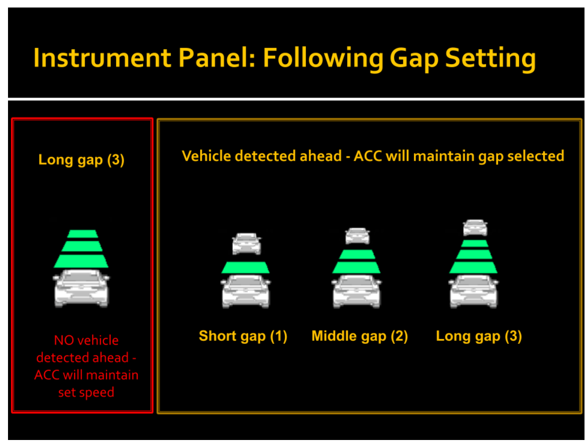


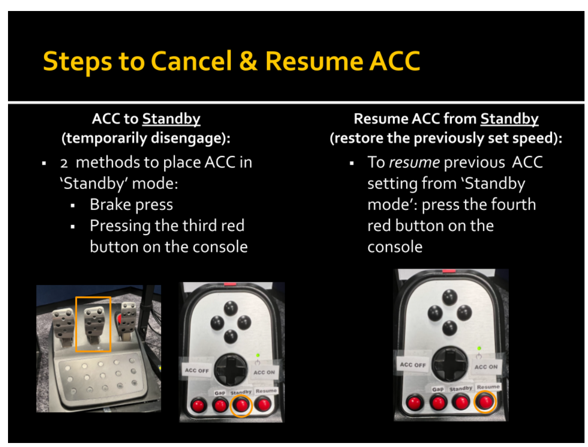




Advanced Group (AAAFTS Site)





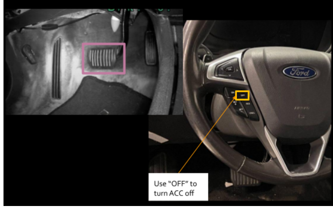


Advanced Group (UMass Site)

Steps to Cancel & Resume ACC

ACC to Off:

- To turn ACC off
  - Press the "OFF" button on the steering wheel
  - Depress the brake pedal



Resume ACC from Off:

- To *resume* ACC from 'Off':
  - Press the "ON" button on the steering wheel (ACC On)
  - Set the ACC speed by pressing the "SET +" or "SET -" button on the steering wheel
  - Press the "OK" button to adjust the following gap if desired

ACC Summary

Speed must be over 30 mph to activate ACC

Steps to Activate ACC

1. Turn ACC 'On-Off': Press the "ON" button on the steering wheel
2. Adjustments to speed: Press the "SET +" or "SET -" button on the steering wheel
3. Set following gap: Press the "OK" button on the steering wheel (one press: short, second press: medium, third press: long)

Cancel & Resume ACC

ACC to 'Standby': press "CNCL" or depress gas

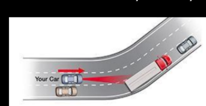
RESUME previous ACC settings: press "RES" or "SET +"

ACC to 'Off': press "OFF" or depress brake

Reactivate ACC: Need to reactivate using the 3 steps listed above

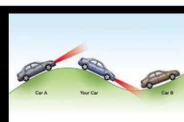
ACC Limitations

ACC uses a radar system to detect vehicles directly in front of you. Radar may identify traffic in another lane and may NOT always be able to detect vehicles in your lane. This occurs because vehicles may or may not be in the radar's field of view.



ACC behavior may respond differently in these and other situations:

- Sharp curves
- Turning vehicles
- Roadway elevation changes/hills
- Extremely slow moving vehicles



Advanced Group (AAAFTS Site)

Steps to Cancel & Resume ACC

ACC to Off:

- To turn ACC off: Press the Left button on the D-Pad

Resume ACC from Off:

- To *resume* ACC from 'Off':
  - Press the Right button on the D-Pad (ACC On)
  - Set the ACC speed by pressing the Up or Down button on the D-Pad
  - Press the second red button on the console to adjust the following gap if desired

ACC Summary

Speed must be over 40 mph to activate ACC

Steps to Activate ACC

1. Turn ACC 'On-Off': Press the Right button on the D-pad
2. Set ACC speed: Press the Up or Down button on the D-pad
3. Adjustments to speed: Press the Up or Down button on the D-Pad in the direction needed
4. Set following gap: Press the second red button on the console (default gap: short)

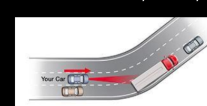
Cancel & Resume ACC

ACC to 'Standby': RESUME previous ACC settings = Press fourth red button (on the console)

ACC to 'Off': Reactivate ACC = Need to RESET the ACC settings using the 4 steps listed above

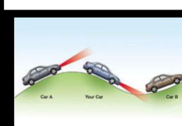
ACC Limitations

ACC uses a radar system to detect vehicles directly in front of you. Radar may identify traffic in another lane and may NOT always be able to detect vehicles in your lane. This occurs because vehicles may or may not be in the radar's field of view.



ACC behavior may respond differently in these and other situations:

- Sharp curves
- Turning vehicles
- Roadway elevation changes/hills
- Extremely slow moving vehicles

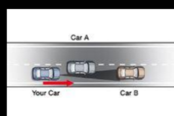


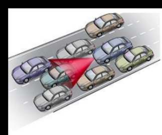
Advanced Group (UMass Site)

ACC Limitations

ACC behavior may respond differently in these and other situations:

- Merging vehicles
- Smaller vehicles like motorcycles
- Vehicles not in the center of the lane ahead
- Heavy traffic with frequent repeated acceleration & deceleration
- Stopped vehicles or objects in the road (construction cone, tire, ball)
- Poor weather conditions (e.g., fog, snow, rain, ice) = covered sensors



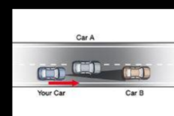


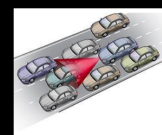
Advanced Group (AAAFTS Site)

ACC Limitations

ACC behavior may respond differently in these and other situations:

- Merging vehicles
- Smaller vehicles like motorcycles
- Vehicles not in the center of the lane ahead
- Heavy traffic with frequent repeated acceleration & deceleration
- Stopped vehicles or objects in the road (construction cone, tire, ball)
- Poor weather conditions (e.g., fog, snow, rain, ice) = covered sensors





Appendix B: Trust Survey (Jian et al., 2000)

Please answer the following questions based on your understanding of Adaptive Cruise Control (ACC): (participants rated on a 7-point scale: 1 (Not at all) – 7 (Extremely))

- The system is deceptive
- The system behaves in an underhanded manner
- I am suspicious of the system's intent, action, or outputs
- I am wary of the system
- The system's action will have a harmful or injurious outcome
- I am confident in the system
- The system provides security
- The system has integrity
- The system is dependable
- The system is reliable
- I can trust the system
- I am familiar with the system

Appendix C: Technology Acceptance Survey (Van der Laan et al., 1997)

The use of the system while driving would be:								
	1	2	3	4	5	6	7	
Bad								Good
Useless								Useful
Desirable								Undesirable
Ineffective								Effective
Sleep inducing								Alerting
Unpleasant								Pleasant
Extremely Annoying								Not at all annoying
Irritating								Likable
Assisting								Worthless

Appendix D: Mental Model Assessment

The following questions will ask you about your current understanding of Adaptive Cruise Control (ACC). You may have previous experience with ACC systems, but it is important that you focus on the information provided to answer these questions. For each question, please indicate whether the statement is "True" or "False". Please answer the following question regarding ACC.

For each item, participants were also asked: “Please indicate your level of confidence to the previous question” (Low confidence, Medium confidence, High confidence)

1) Maintains the speed that you have set when there are no vehicles detected in the lane ahead

2) Brakes and accelerates to maintain a following gap from the vehicle ahead

3) Adjusts the speed to match faster vehicles ahead

4) Will accelerate if a slower vehicle ahead moves out of the detection zone

5) Will provide steering input to keep the vehicle in its lane

6) Will correctly detect motorcycles and other smaller vehicles not driving in the center of the lane

7) Is meant to be used on highways and interstates

8) May not correctly detect stopped vehicles in your lane

9) Reacts to stationary objects on the road (construction cone, tire, ball)

10) Works well on curvy roads and hills

11) Is meant to be used in slow and heavy traffic

12) Adjusts the speed when there are slower moving vehicles detected ahead

13) Will react immediately to vehicles merging onto the road in front of you

14) Reacts to oncoming traffic

15) Adjusts the vehicle speed when approaching tight curves

16) Is meant to be used on rural roads

17) May not correctly detect vehicles ahead traveling at much slower speeds

18) Works even when the radar sensor is dirty

19) Can be activated at a standstill

20) Can handle operating in all weather conditions