DEVELOPING A STANDARDIZED PERFORMANCE EVALUATION OF VEHICLES WITH AUTOMATED DRIVING FEATURES

ABSTRACT

Objectives: *The project goal was to create an initial set of standardized tests to explore whether they enable the ongoing evaluation of automated driving features as they evolve over time. These tests focused on situations that were representative of several daily driving scenarios as encountered by lower-level automated features, often called Advanced Driver Assistance Systems (ADAS), while looking forward to higher levels of automation as new systems are deployed.*

Methods: *The research project initially gathered information through a review of existing literature about ADASs and current test procedures. Thereafter, a focus group of industry experts was convened for additional insights and feedback. With this background, the research team developed a series of tests designed to evaluate a variety of automated driving features in currently available implementations and anticipated future variants. Key ADAS available on current production vehicles include adaptive cruise control (ACC), lane keep assist (LKA), and automatic emergency braking (AEB). Seven of the most automated production vehicles available in 2018 from six manufacturers were subjected to a series of standardized tests that were performed on a closed test track environment to assess the vehicle capabilities and limitations of the automated driving systems' operational domains.*

Results: *Considerable performance variability was observed between different vehicle manufacturers and within a single vehicle model across repeated trials and multiple replications. In addition, there were specific roadway characteristics that significantly impacted performance.*

Conclusions: *The results indicate that standardized testing can assist researchers in determining the current capabilities of vehicles with automated driving features. The research team suggests continuing to improve and expand standardized testing of automated driving features and to work toward industry consensus of a robust evaluation mechanism that may play a key role in the conformance of future automated-vehicle systems.*

Keywords: Automated vehicles, advanced driver assistance systems, evaluations, ACC, LKA, AEB

INTRODUCTION

Vehicles with automated driving features are becoming more prevalent as manufacturers and consumers understand the need for safer and more convenient modes of transportation. However, not all automated features are the same. There are different levels of automated system capabilities, ranging from readily available vehicles with Advanced Driver Assistance Systems (ADAS) through the envisioned, fully automated vehicles of the future. The Society of Automotive Engineers (SAE) has published a Recommended Practice that categorizes these levels of automation within vehicles [1]. Although higher levels of automation for vehicles are on the horizon, currently the most widely available levels of automation are Level 1 (L1) and Level 2 (L2). If appropriately equipped, these vehicles offer driver assistance and active safety features, yet still require full-time driver engagement and monitoring of the systems and driving environment. Features of interest in these production vehicles include those that automate lateral and longitudinal control (e.g., adaptive cruise control [ACC], automatic emergency braking [AEB], and lane keep assist [LKA]). Although these systems fall under the same L2 distinction, ACC and LKA features are engaged manually by the driver and provide continuous vehicle control until they are disengaged. AEB systems are available without driver input, triggering when the system senses hard braking is required to avoid an impending collision.

The near-term goal of this project was to develop and evaluate an initial set of standardized test procedures that vehicles equipped with automated driving features could undergo to compare capabilities and limitations across different implementations of automated technologies. In the longer term, this project will provide a basis for future automated-vehicle testing and gather evidence to help researchers determine the value of standardized testing as part of an automated-vehicle conformance process. Using Virginia Tech Transportation Institute (VTTI) vehicles equipped with automated driving features (i.e., ACC, LKA, and AEB), this initial project focused on answering two research questions:

1. What testing should be conducted to evaluate both current and forthcoming capabilities of vehicles with automated driving features?
2. How do currently available vehicles with automated driving features perform under the proposed standardized set of evaluations?

Minimal previous research has been reported in the area of automated feature testing, with most of the research focusing on theory or examining extreme edge-case scenarios [2, 3, 4]. In addition, emphasis has been placed on testing and developing standardized testing protocols for singular automated systems but not a broader automated feature set [5, 6, 7]. Until recently, there have been no previous studies evaluating the capabilities of current automated driving features in routine, daily driving scenarios [8, 9].

METHODS

The following section summarizes the methods employed during this project, with a focus on the scenarios under evaluation. Scenarios were inspired by literature reviews, real-world driving events gathered through reviews of media reports, and current experimentation conducted with these types of vehicles [10]. Through literature and media reviews, the research team identified current complexities of automated feature testing, such as the need for elaborate test setups [11], uncertainties and variability of real-world driving [12], safety concerns associated with physical vehicle testing [13], and real-world scenarios in which vehicles with automated features typically fail or experience difficulty. Vehicle owner's manuals were consulted during development to gain better understanding of the operational design domain (ODD) of specific automated driving features. Test procedures were designed and theoretically validated in collaboration with a focus group of automated- and connected-vehicle experts at VTTI, where safety-critical automated systems were highlighted, current automated-vehicle testing practices were discussed, and preliminary testing scenarios were reviewed.

Testing of finalized scenarios occurred on the Virginia Smart Roads Highway and Surface Street sections, both of which are closed test tracks located at VTTI. The Highway portion of the Smart Roads is approximately 2.2 miles long and built to Federal Highway Administration specifications. The road features multiple lanes, two high-speed turnarounds, and numerous opportunities for customization to create various real-world scenarios. The Surface Street is a multi-use test facility that specifically allows for more city-like driving, with multiple turns, intersections, and turnaround points. This section also has adjustable lane markings, allowing complete customization for replication of a wide array of urban scenarios.

Testing involved varying relevant factors in each scenario (e.g., speed, obstruction placement, size of obstacle) to cover multiple use cases for automated driving features. Three replications of each trial were conducted to evaluate the consistency of system performance. To minimize variance due to operator factors, tests were performed using four highly skilled drivers, each with experience testing advanced-vehicle features. These drivers were given instructions prior to each test regarding execution strategies and takeover procedures; these instructions can be found in Appendix A. Although these procedures were established to increase safety and minimize variation, the study drivers had ample experience conducting advanced-vehicle tests with automated features. Therefore, decisions to swerve, brake, or abort a trial were ultimately made at their discretion.

Study Vehicles

To assess the value of standardized testing, a variety of on-road tests were developed for a fleet of vehicles currently available on the market and equipped with automated driving features. A summary outlining the study vehicles used and their associated automated technology packages can be found in Table 1. The ODD speeds for each feature tested can be found in Appendix B.

Table 1: Vehicles Used in the Study

Vehicle Make/Model	**Automated System Package Included in Vehicle**
2015 Infiniti Q50 3.7 AWD Premium	Technology, Navigation, and Deluxe Touring
2015 Tesla Model S P90D AWD	Autopilot Convenience
2016 Mercedes-Benz E350 Sedan	Premium Package, Driver Assistance
2016 Volvo XC90 T6 AWD R-Design	Convenience
2017 Audi Q7 Premium Plus 3.0 TFSI Quattro	Driver Assistance

2018 Cadillac CT6 AWD 3.6L Engine Premium Luxury	Driver Awareness and Convenience, Super Cruise
2018 Tesla Model X P100D AWD	Autopilot Convenience

Test Procedures and Scenarios

Detailed documentation was captured by researchers for each test, recording vehicle behavior, test parameters, and general observations regarding performance of the automated driving features. All drivers maintained communication throughout the entirety of tests via two-way radios. In addition, objective performance variables such as speed, acceleration, and GPS location were collected using VTTI data acquisition systems (DAS) installed in the test vehicles. Detailed analyses of these additional data sources are to be performed in a future activity. Scenarios are described in the section below, with the parameters varied for each trial shown in Table 2. All testing used a full factorial design method to vary parameters across scenarios.

Table 2: Test Parameters

ACC Curve					
Curve Radius (ft)	205	108	301	295	
Lead Vehicle Speed (mph)	15	20	25		
Test Vehicle ACC Set Speed (mph)	20	25	30		
Headway Setting*	Long	Medium			
ACC Cut-In					
Lead Vehicle Speed (mph)	30	40	50		
Test Vehicle Speed (mph)	35	45	55		
Cut-In Vehicle Speed (mph)	30	40	50		
Headway Setting*	Long	Medium			
ACC Cut-Out & Reveal					
Cut-Out Vehicle Speed	20	30	40		
Test Vehicle Speed	25	35	45		
Revealed Vehicle Speed	0	10	15		
Headway Setting*	Long	Medium			
ACC Stop & Go					
Headway Setting*	Long	Medium			
AEB Obstacle					
Test Vehicle Speed (mph)	25	35	45		
Obstacle Type	Static large pedestrian	Static small pedestrian	Dynamic large pedestrian	Dynamic small pedestrian	Foam car
Lane Obstruction					
Test Vehicle Speed (mph)	25	35	45		
Obstacle Type	Foam car	Barrel cone			
Headway Setting*	Long	Medium	Short		
Lane Shift					
Test Vehicle Speed (mph)	25	35	45		
Lane Shift Severity	Half lane width	Full lane width			
Headway Setting*	Long	Medium	Short		

*Attempts were also made to reduce variability between the different headway settings of the vehicles, which, within the fleet, ranged from 3-7 different modes.
Through vehicle manuals, research, and physical testing, each mode's following distance time was determined, and three variations—long (following distance of ~3s), medium (following distance of ~2s), and short (following distance of ~1s)—were identified for use during testing.

The following test scenarios were evaluated:

ACC Curve: The ACC Curve test evaluated the effectiveness of ACC and LKA features when exposed to curved roadways of varying radii. In this evaluation, the test vehicles were set to a specified ACC speed (i.e., 5 mph higher than lead vehicle speed to maintain coupling) and headway while following a lead vehicle traveling at a constant speed around curves of four different radii and geometries on the Highway portion of the Smart Roads, as seen in Figure 1.

Figure 1: ACC Curve scenario diagram.

Researchers examined whether the test vehicle tracked the lead vehicle throughout the entire curve either by visual indicators provided on the factory human-machine interface (HMI) display, through surges in test vehicle speed, or if the test vehicles ceased tracking the lead vehicle and instead began to accelerate, requiring the test vehicle driver to intervene.

ACC Cut-In: The ACC Cut-In test was used to evaluate how each ACC implementation reacted to a new vehicle merging into the existing gap between the test and lead vehicles. This test was conducted on approximately one mile of straight road on the Highway portion of the Smart Roads. During the procedure, as seen in Figure 2, the ACC of each test vehicle was set to a specific speed and headway following a lead vehicle, while a second vehicle ("cut-in vehicle") traveled in the adjacent lane in the same direction. Once the lead and test vehicles reached steady state (i.e., both were holding the constant set speed), the cut-in vehicle began to merge between the test vehicle and the lead vehicle.

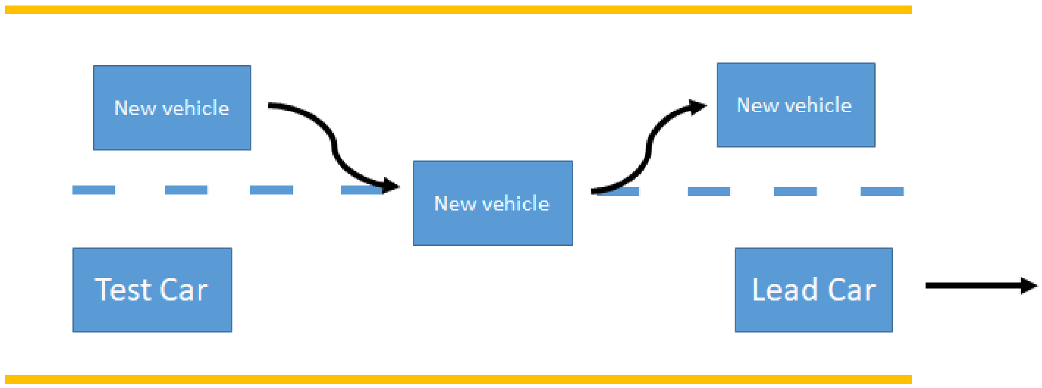

Figure 2: ACC Cut-In scenario diagram.

The cut-in vehicle held a constant speed while centered over the lane line, allowing trained experimenters in the test vehicles to examine the behavior of the test vehicles with the addition of the cut-in vehicle.

ACC Cut-Out and Reveal: The ACC Cut-Out test was used to evaluate the reaction of each test vehicle to a situation where the established lead vehicle changed lanes, revealing a new lead vehicle that, in some cases, was traveling at a drastically slower speed. The original lead vehicle (the "cut-out" vehicle) was followed by the test vehicle that was set to a specified ACC speed (i.e., 5 mph higher than the lead vehicle speed) and headway. Approximately 0.25 miles ahead of the lead and test vehicles, another vehicle (the "revealed" vehicle) was traveling at a specified speed or stationary. The test vehicle and cut-out vehicle traveled toward the revealed vehicle in the same lane; when the pair was approaching, the lead vehicle changed lanes when the time-to-collision (TTC) between the lead vehicle and reveal vehicle was approximately 3s, thus revealing the slower moving vehicle, demonstrated in Figure 3.

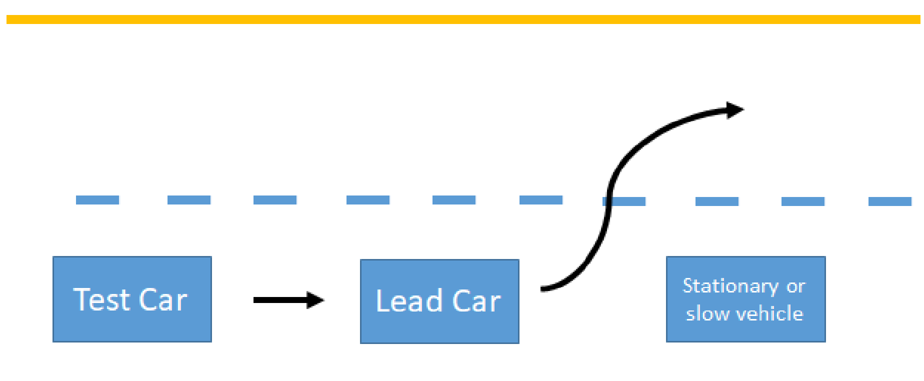

Figure 3: ACC Cut-Out and Reveal scenario diagram.

Researchers made detailed notes of how each ACC implementation reacted to this newly revealed, slower-moving or stopped lead vehicle.

ACC Stop and Go: The ACC Stop and Go test was used to examine the ability of each test vehicle to react to a lead vehicle that was frequently changing speeds. The scenario, shown in Figure 4, begun with the test vehicle positioned behind two lead vehicles on the Highway section of the Smart Roads, with ACC engaged and set to a speed of 35 mph. At the start of the test, the lead vehicles both accelerated to 30 mph while maintaining a 2s following distance between each other. Once the lead vehicles reached steady state, an in-vehicle researcher cued the lead vehicle drivers to decelerate to 15 mph. Once all vehicles reached steady state after deceleration, the lead vehicle drivers accelerated to 25 mph until steady state was again achieved. Finally, the lead vehicle drivers decelerated to a stop. After all vehicles were stationary for 3s, the lead vehicle drivers accelerated back to 30 mph, with the test vehicle following. All acceleration and deceleration performed by the drivers were representative of levels anticipated in typical stop-and-go traffic, or approximately 0.2g - 0.3g [14].

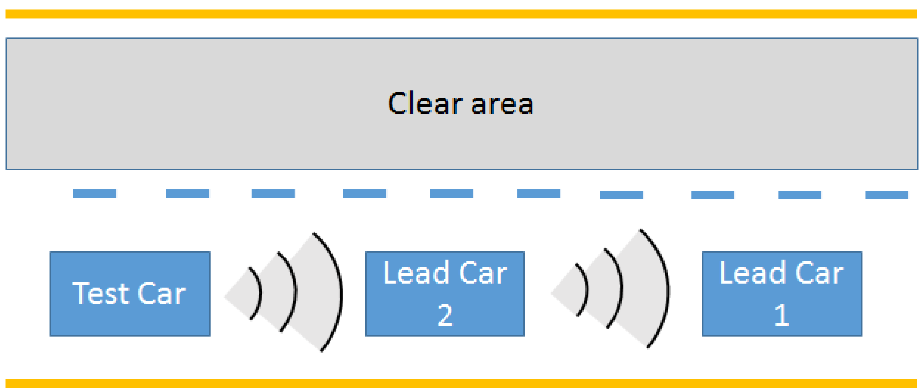

Figure 4: ACC Stop and Go scenario diagram.

Researchers took note of how each ACC implementation performed in terms of matching the changing speeds and adjusting following distances once regaining an appropriate following position.

LKA Inattentiveness: The LKA Inattentiveness test evaluated the reaction of the test vehicle to a driver taking his or her hands off of the steering wheel while ACC and LKA were engaged. At specified points along the road, drivers would remove their hands from the steering wheel and hover them over the wheel as the vehicle maintained lane centering and speed. Researchers recorded how long it took the vehicle to react to the lack of driver input and what types of warnings were presented (i.e., visual, auditory, haptic). If the ACC disengaged automatically due to the lack of driver input, the test driver was instructed to allow the vehicle to drift out of the lane by one car’s width, to determine if any sort of preventative measures to resist the lane drift would be taken by the vehicle. However, since observations were more informative and not performance-based, the results from this scenario are not discussed in further detail.

AEB Obstacle: The aim of this test was to determine the AEB capabilities of each test vehicle when presented with a variety of soft-target objects in the driving path.

Expert drivers drove the test vehicles at a specified speed toward a stationary or dynamic obstacle, seen in Figure 5, which was set up directly in (static) or moved into (dynamic) the center of the traveling lane by an on-road researcher. The dynamic obstacles were controlled by an additional on-road researcher using a robotic, remote evasive maneuvering device (i.e., a small robotic platform constructed by VTTI).

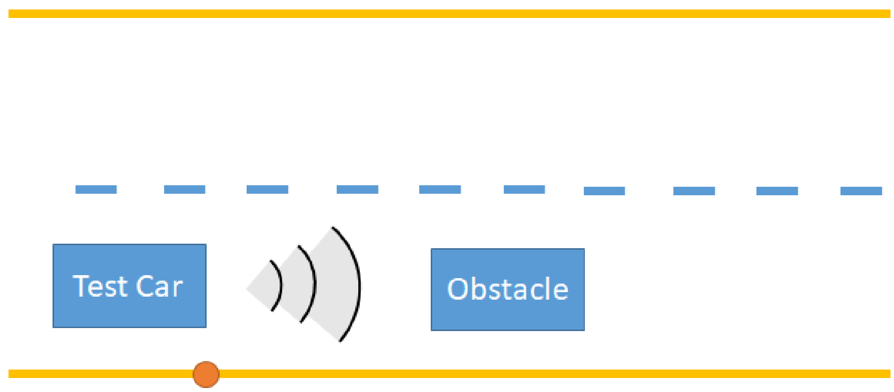

Figure 5: AEB Obstacle scenario diagram.

When approaching the obstacle at the specified trial speed, the driver was prepared to perform an evasive maneuver in case the system did not recognize the hazard and brake accordingly. Researchers observed whether or not the vehicle reacted to the obstacle in the driving path, either by indication of the dashboard HMI display or by the application of automatic braking.

Lane Obstruction: Similar to the AEB Obstacle test, this test aimed to determine the ability of each test vehicle to recognize stationary obstacles in the driving path and react accordingly. However, in this case, the ACC capabilities of each test vehicle were also examined to determine if the vehicle would attempt to mitigate or avoid the obstacle.

In this test, the obstacle was only partway in the lane rather than fully centered, shown in Figure 6, simulating an obstacle on the shoulder of a road protruding into the traveling lane. A single standard barrel cone or strikable soft car was used as stationary obstacles for this test. Test drivers drove toward the obstacle at a specified ACC speed and headway setting. When approaching the obstacle while maintaining the desired speed for each trial, the driver was prepared to perform an evasive maneuver in case the system did not recognize the hazard and brake accordingly. Similarly to the AEB Obstacle test, researchers recorded whether the vehicle reacted to the obstacle in the driving path either by the indication on the dashboard HMI display or by the adjustment of speed by ACC.

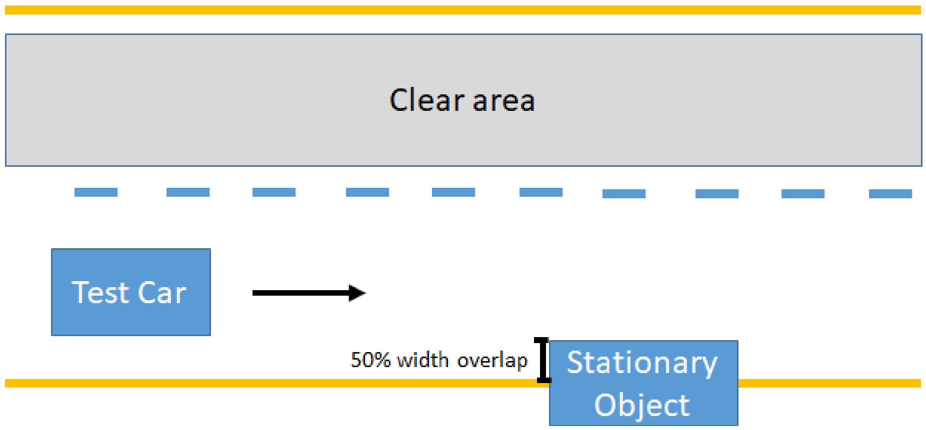

Figure 6: Lane Obstruction scenario diagram.

Lane Shift: This test was inspired by several videos that were publicly released prior to the current study; these videos showed vehicles with lower levels of automated driving features colliding into temporary barriers set up in construction zones indicating lane shifts. The test examined the ACC system capabilities of each test vehicle in a barrel cone barrier lane-shift scenario. Both full- and half-lane shifts across a standard ~10-foot lane were set up according to standards [15]. Cones were placed approximately 10 feet apart to create a smooth lane transition typical of work zones. The setup for this scenario can be seen in Figure 7 below.

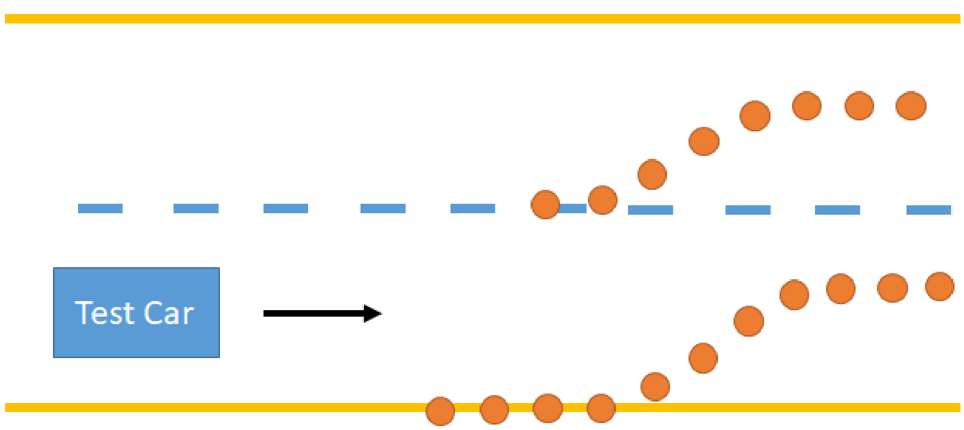

Figure 7: Lane Shift scenario diagram.

Data Analysis

As the primary objective of this study was to develop draft standardized testing procedures that could be replicated with minimal to no vehicle instrumentation, data obtained and analyzed from testing were largely qualitative. Narratives of vehicle reactions and general observations about tests were recorded by in-vehicle documenters during trials. To convert the qualitative results into quantitative data for analysis, coding schemes were developed based on vehicle performance in each test. These coding schemes categorized whether the vehicle had no reaction, issued auditory/visual alerts, attempted mitigation, or demonstrated complete avoidance and/or expected reactions to different obstacles or a changing driving landscape, as seen in Appendix C. It is important to note that these codes are based on the performance of the vehicles within the specific tests and are not directly related to regulations, standards, or crash data. When possible, future work should attempt to refine the connection between performance surrogate measures, such as those captured herein, and crash data for automated driving features.

Referencing the specific code scheme assigned to each test, two researchers independently assigned codes to individual results of each test trial. As needed, video captured during the experiments was reviewed to confirm the observations made during the tests. By using inter-rater reliability methods, it was found that, in general, the researchers' results were highly consistent; occasional discrepancies in code assignments between both researchers were discussed and resolved. If a resolution was not achieved between the coding researchers, additional team members were consulted until a final decision was made.

Based on the quantitative data obtained, average performance scores for each test vehicle and for each test replication were derived by computing the mean score for each test trial. These scores represent the levels at which the vehicles tended to identify and respond to the variations of stimuli during testing. The higher the performance score, the more consistently vehicles exhibited appropriate behavior. Full details of these scores are available in Appendix D. The average performance scores for each vehicle, across all executed trials, provided the basis for the analysis of variance (ANOVA) testing.

Single-factor ANOVA tests with an alpha value of 0.05 were performed to determine statistically significant differences in performance across the examined feature implementations. If the ANOVA indicated that there was statistical significance, two tailed t-tests assuming equal variances were performed to determine which variable had the most significant impact on vehicle performance. Statistical analysis results, including ANOVAs and t-tests, can be found in Appendix E.

RESULTS

To reduce bias during data analysis and to protect manufacturer identity in publication, vehicle make and model information were removed and were represented numerically in the results. The testing generated a sizeable data set; therefore, the results presented below represent only a snapshot of the total data obtained through testing. However, the results presented herein demonstrate the greatest differentiation between vehicle capabilities. Complete results will be available through the Safety through Disruption National University Transportation Center Report #VTTI 00-20.

Curve Radius Influenced Test Vehicle Following Performance while ACC was Engaged

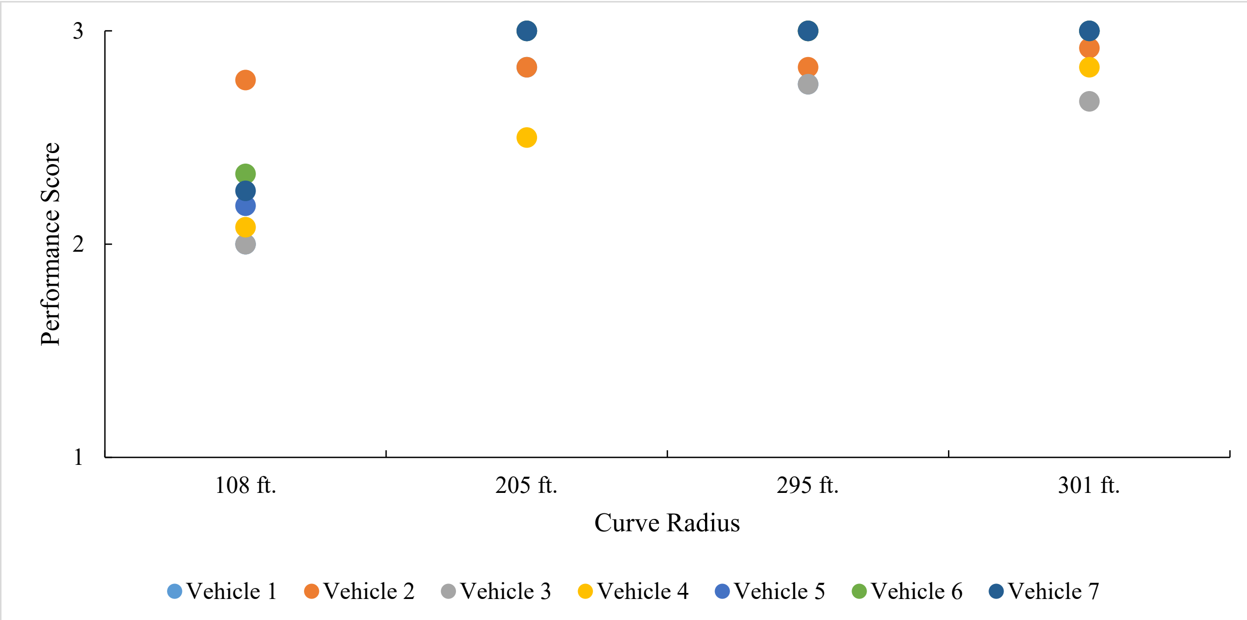

Figure 8: Average vehicle performance during ACC Curve test (out of a maximum performance score of 3), considering curve radius.

Statistical analyses indicated that the different curve radii had significant impacts on vehicle performance. With one exception, ACC implementations performed slightly worse than what was seen in the overall averages while traveling around the tightest curve radius (108-foot radius); this was the only curve where driver intervention was needed. Since the angle of the curve was tighter than the other curves, it is hypothesized that typical sensor and camera field of views were unable to track the lead vehicle throughout its duration [16, 17].

Headway Setting Affected Test Vehicle Following Performance while ACC was Engaged

Although only by a small margin, all ACC implementations consistently performed better with the medium headway setting engaged versus a long headway. This performance difference is seen particularly in the ACC Curve test, where the different headway settings produced statistically significant results. A medium ACC headway setting allowed the test vehicle to follow the lead vehicle more closely, relative to a long headway. As such, the lead vehicle stayed in the sensor and camera field of view for longer around curves and often throughout the full curve traversal, thus improving performance [16, 17]. Due to safety restrictions, the lowest headway setting was not tested. However, it is hypothesized that a trend of improved performance due to a closer following distance would likely continue.

Test Vehicle ACC Performance was Affected when a Revealed Vehicle was Stationary

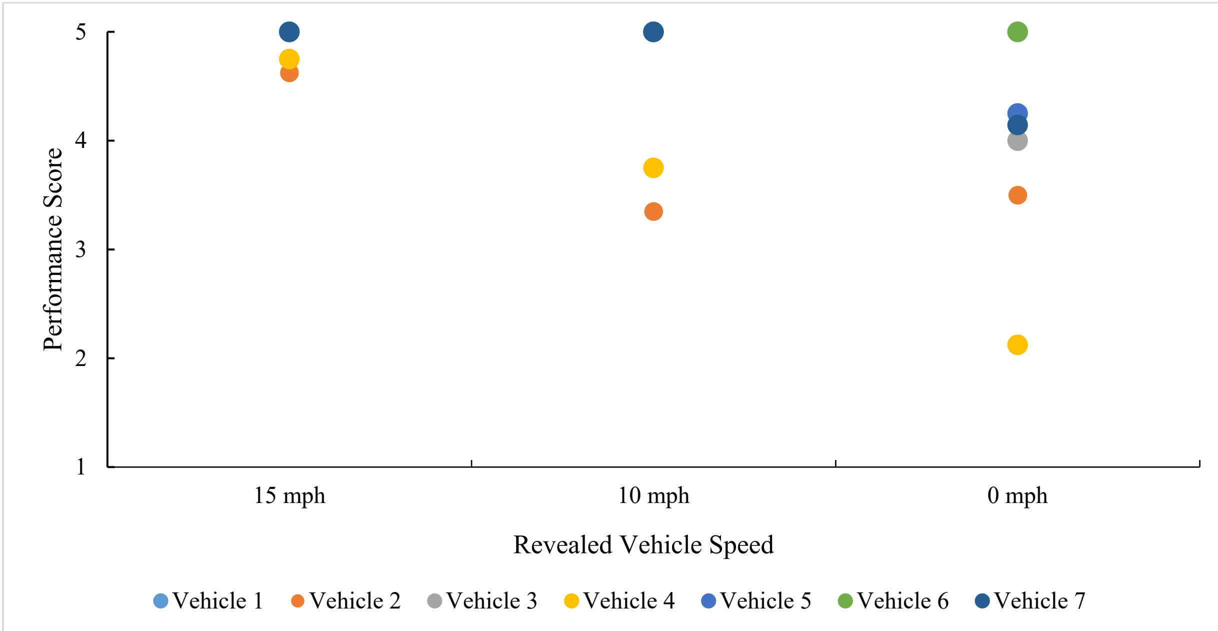

Figure 9: Vehicle performance scores (out of a maximum score of 5) when exposed to different revealed vehicle speeds during the ACC Cut-Out & Reveal test.

While examining the ACC Cut-Out test data, the only t-tests that produced significant results were for scenarios wherein the revealed vehicle was stationary (0 mph), specifically when the test vehicles were approaching at 30 mph. Therefore, it can be concluded that the revealed vehicle speed was the main factor that influenced test vehicle performance during the cut-out/reveal events. Qualitatively, in trials where the revealed vehicle was traveling at speeds of 15 mph and 10 mph, test vehicles performed in alignment with researchers' initial expectations by consistently recognizing and responding to the new lead vehicle. Although trials where the revealed vehicle was stationary had low performance scores relative to the other trials, all test vehicles still responded to the obstacle in the driving path. A likely explanation for these observations is that current ACC is not specifically designed to handle larger speed delta situations, particularly those involving stopped and newly acquired lead vehicles. This type of scenario would be better suited for an evaluation of AEB systems, which some test vehicles initiated before driver intervention.

Test Vehicles Were Unable to Identify Barrel Cones as Obstacles

Although barrel cones have a sharp color contrast against the background environment and reflectivity strips, which should theoretically make them an easy target to identify by camera and some radar solutions, none of the test vehicles reacted to them during the Lane Obstruction test, and nearly none identified them during the Lane Shift test. This response absence may reflect the current ACC capabilities of the vehicles, which may not be specifically designed to handle situations with stationary objects or work zone barriers. These results indicate that none of the current automated perception systems are equipped to detect smaller, potentially safety-critical obstacles, which could pose threats especially in work zones or emergency situations – a key requirement for high levels of future automated systems.

Test Vehicles Exhibited Inconsistent Response Type and Frequency to Identical Conditions

Test vehicles demonstrated inconsistent performance both across the vehicle fleet and within a single vehicle across multiple trials. Poor vehicle performance was often noted during tests where stationary obstacles were present. As mentioned previously, limitations of current ACC systems may contribute to this degraded performance. However, all test vehicles were equipped with AEB systems, so it is surprising, if not a bit troublesome, that most of them did not react well to these targets at close proximity, thus requiring evasive steering input from expert drivers. One possible explanation for the poor response could be the targets that were selected, since both the dynamic pedestrian and foam car targets have not undergone extensive validation. However, both objects were recognized as hazards by some vehicles.

Test vehicles also had trouble responding appropriately to the ACC Cut-In test. For some vehicles, the test vehicle did not detect the merging vehicle and would not adjust speed or following distance accordingly. Many ACC system limits are exceeded by this scenario, and this variability could also point to potential issues with limited field of view of both the cameras and sensors on test vehicles. However, the fact that some vehicles performed well under this condition should be noted, although poor performance should be expected in general.

In the LKA Inattentiveness test, the time it took for auditory and/or visual warnings to appear after the driver removed his or her hands ranged widely across vehicles. Some vehicles consistently warned drivers at a specific time interval; for other vehicles, the warning time varied, and for one vehicle, there was no warning before ACC and LKA disengagement.

DISCUSSION

This work supports a growing body of knowledge toward creating standardized tests to evaluate automated driving features, from the ADAS of today into the highly automated vehicle of tomorrow. In addition, this work represents one of the first sets of real-world testing scenarios for evaluation of vehicles with automated driving features, where prior work has focused on smaller scale testing, simpler testing, or testing in virtual environments.

Across the results presented, researchers made the following general observations:

- Test vehicles exhibited expected and more consistent performance in scenarios that simulated higher speeds (i.e., highway scenarios rather than lower speed urban scenarios).
- Many of the test vehicles exhibited a response (e.g., warning lights/sounds) to obstacles in their path but did not always implement a crash avoidance strategy.
- Large inconsistencies in test vehicle performance occurred within a vehicle platform across repeated trials and across the fleet evaluated.

The findings are important because they demonstrate that vehicles with currently available automated driving features exhibit large variations in capabilities. Additionally, results indicate control scenario testing can provide means to evaluate improvements in automated capabilities as systems are refined (e.g., measures should indicate more favorable and consistent outcomes as systems improve). For vehicles with automated features to be trusted and accepted by users, and to leverage these features as building blocks for higher levels of automation, it is critical that their performance be consistent and predictable.

The high variability in vehicle performance and poor interactions with certain test conditions could be due to the automated systems' ODD limitations. For example, current ACC and LKA systems are not typically designed to interact with stationary objects, even though it is a possibility in real-world driving. In addition, certain obstacles used during testing may not be within the systems' machine-vision capabilities. Current ADASs also expect the driver to be engaged in the driving task at all times. In this study, although the driver was focused on the road during every test, he or she was instructed to not perform any sort of corrective maneuver or intervention until ~3s TTC, which may have resulted in the appearance of poorer performance by the vehicles as they are not designed to completely take over driving tasks. This provides a vehicle performance benchmark of currently available systems, with the expectation that future implementations will offer continuous improvement across these scenarios.

Through this project, a framework of standardized tests was developed to evaluate the capabilities of automated driving features in real-world scenarios. In addition to the basic testing framework, researchers were able to use these tests to create baseline results for current production vehicles and to gain a better understanding of the discrepancies and variations of features. Due to the high variability witnessed within tests, this study also helps confirm the need for standardized testing of automated driving features across all levels. Most importantly, this study provides evidence that structured scenario testing of automated driving features may be used to detect differences between vehicle capabilities. While the authors believe considerably more work is needed to develop a robust set of industry standardized testing, this study provides a base strategy for simple systematic evaluation of automated driving features as performance improves over time. Such controlled testing may be part of a future conformance testing process when paired with additional robust methods.

Overall, the work performed herein clearly indicates support for larger, more comprehensive research into the expansion and refinement of standardized test procedures for automated driving features. It also indicates that observational methods that require minimal vehicle instrumentation will provide value toward assessing automated driving features. Such observational methods could be bolstered by additional objective measures collected with higher-precision data recording devices.

DATA AVAILABILITY STATEMENT

Raw data, both qualitative and quantitative, are available upon request. Vehicles will remain coded within the raw data set for privacy reasons. Overall raw scores can be found in Appendix D.

REFERENCES

[1] SAE International. J3016: Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. 2018.

[2] Knauss A, Schröder, J, Berger C, Eriksson, H. Paving the roadway for safety of automated vehicles: An empirical study on testing challenges. 2017 IEEE Intelligent Vehicles Symposium (IV), Los Angeles, CA: 1873-1880.

[3] Zhao D, Peng H. From the Lab to the Street: Solving the Challenge of Accelerating Automated Vehicle Testing. University of Michigan, Mcity White Paper. 2017.

[4] Sippl C, Bock F, Wittmann D, Altinger H, German R. From Simulation Data to Test Cases for Fully Automated Driving and ADAS. 2016 IFIP International Conference on Testing Software and Systems, Graz, Austria: 191-206.

[5] Insurance Institute for Highway Safety (IIHS). Autonomous Emergency Braking (AEB) Test Protocol (Version I). 2013.

[6] Euro NCAP. Test Protocol – AEB systems. 2017.

[7] UNECE. Regulation No. 79: Uniform Provisions Concerning the Approval of Vehicles with Regard to Steering Equipment. 2005.

[8] Insurance Institute for Highway Safety (IIHS). Reality check: Research, deadly crashes show need for caution on road to full autonomy. Special Issue: Autonomous Vehicles; 53(4):1-11. 2018.

[9] Euro NCAP. 2018 Automated Driving Tests. 2018.

[10] Terry T, Tanner S. Problematic Roadway Environments for Automated Vehicles – Database Development. National Surface Transportation Safety Center for Excellence*,* Report #18-UR-064. 2017.

[11] Agaram V, Barickman F, Fahrenkrog F, Griffor E, Muharemovic I, Peng H, Salinger J, Shladover S, Shogren W. Validation and Verification of Automated Road Vehicles. Road Vehicle Automation 3: 201-210. 2016.

[12] Pfeffer R, Leichsenring T. Continuous Development of Highly Automated Driving Functions with Vehicle-in-the-Loop Using the Example of Euro NCAP Scenarios. Simulation and Testing for Vehicle Technology: 33-42. 2016.

[13] Alberi T. A Proposed Standardized Testing Procedures for Autonomous Ground Vehicles. Virginia Tech Master's Thesis. 2008.

[14] Koppa R. Monograph on Traffic Flow Theory. U.S. Department of Transportation, Federal Highway Administration: 1-32. 1990.

[15] Virginia Department of Transportation. Work Zone Safety: Guidelines for Temporary Traffic Control. 2011 Virginia Work Area Protection Manual: 14-15. 2012.

[16] International Organization for Standardization (ISO). ISO 22179: Intelligent transport systems- Full speed range adaptive cruise control (FSRA) systems- Performance requirements and test procedures. 2009.

[17] International Organization for Standardization (ISO). ISO 15622- Intelligent transport systems- Adaptive Cruise Control systems- Performance requirements and test procedures. 2010.

APPENDICES

A. Driver Performance Expectations

Test	Driver Expectations
ACC Curve	• During curves, pay close attention to the vehicle path and any sensors that are tracking the lead vehicle. • Monitor lead vehicle sensing display (if available) and make note if the test vehicle ceases to track the lead vehicle at any point during the curve. • Keep hands on the wheel and prepare to take over if the vehicle begins to deviate from the curve path. • Swerve or brake when time to collision (TTC) is ~3s.
ACC Cut-In	• Pay close attention to how the vehicle reacts to the cut-in vehicle. If test vehicle does not sense the cut-in vehicle, be prepared to brake or correct steering. • Monitor lead vehicle sensing display (if available) and make note if the test vehicle recognizes the new lead vehicle. • Keep hands on the wheel and prepare to brake or take over if the test vehicle gets too close to the lead vehicle. • Swerve or brake when time to collision (TTC) is ~3s.
ACC Cut-Out & Reveal	• Pay close attention to how the vehicle reacts to revealed vehicle. If test vehicle does not sense the revealed vehicle, be prepared to brake or swerve to avoid collision. • Prepare to brake hard or swerve in the event that the revealed vehicle is not sensed. • Monitor speed and following distance and intervene if they deviate from set parameters. • Swerve or brake when time to collision (TTC) is ~3s.

ACC Stop & Go	• Pay close attention to how the vehicle reacts to the changing lead vehicle.
	• If test vehicle does not sense the change of speed, be prepared to brake or swerve to avoid collision.
	• Driver intervention needed if:
	  o Test vehicle stops following lead vehicles and disengages ACC
	  o Test vehicle does not brake within a safe distance of the lead vehicles
	  o Test vehicle does not slow down when approaching a slower moving/stopped vehicle
	• Swerve or brake when time to collision (TTC) is ~3s.

AEB Obstacle	• If test vehicle does not sense the obstacle in the driving path, be prepared to brake hard.
	• Remember that there is a researcher on the side of the road controlling the HV-REMO device:
	  o If swerving is necessary, swerve away from the shoulder
	  o Try to brake instead of swerve
	• Swerve or brake when time to collision (TTC) is ~2s.

Lane Obstruction	• Driver be ready to perform an evasive maneuver in the event the vehicle fails to react to the obstacle. • Pay close attention to how the vehicle reacts to the obstacle. • If vehicle fails to detect object, prepare to engage brakes or swerve to adjacent lane to avoid it. • Swerve or brake when time to collision (TTC) is ~2s.
Lane Shift	• Pay close attention to how the vehicle reacts to the lane shift. • Approach cones of lane shift ready to swerve or brake. • Driver needs to intervene if the test vehicle fails to recognize the cones and is not able to recognize them even in correct positioning. • If there is an on-road researcher present and swerving is necessary, swerve away from the shoulder. Try to brake instead of swerve.

B. Operational Design Domain (ODD) of ADAS Features

Vehicle	Adaptive Cruise Control ODD	Lane Keeping Assist ODD	Automatic Emergency Braking ODD (Moving objects)	Automatic Emergency Braking ODD (Stationary objects)
2015 Infiniti Q50	20 mph-90 mph	45 mph-90 mph	0 mph-45 mph	0 mph-45 mph
2015 Tesla Model S	18 mph-90 mph	18 mph-90 mph	7 mph-90 mph	7 mph-90 mph
2016 Mercedes-Benz E350	0 mph-125 mph	0 mph-125 mph	18 mph-155 mph	18 mph-45 mph
2016 Volvo XC90	10 mph-80 mph	40 mph-90 mph	0 mph-50 mph	0 mph-50 mph
2017 Audi Q7	20 mph-95 mph	40 mph-90 mph	20 mph-40 mph	20 mph-40 mph
2018 Cadillac CT6	16 mph-90 mph	37 mph-90 mph	2 mph-50 mph	2 mph-50 mph
2018 Tesla Model X	18 mph-90 mph	18 mph-90 mph	7 mph-90 mph	7 mph-90 mph

C. Coding Scheme Examples

ACC Curve, ACC Cut-In, ACC Stop and Go: Results Key		
Value	**Description**	**Example**
3	No driver intervention needed No false positives or failure of system	Test vehicle adjusts following distance/brakes according to lead vehicle speed.
2	Intermittent system failure No driver intervention needed	Test vehicle briefly/intermittently loses track of lead vehicle but regains tracking before the vehicle begins to deviate from the curve or .decreases following distance drastically
1	Driver required to input lateral or longitudinal control to avoid collision	Test vehicle loses track of lead vehicle completely and does not regain tracking. Driver must brake or correct steering to complete the test.

ACC Cut-Out & Reveal, AEB Obstacle, Lane Obstruction, Lane Shift: Results Key		
Value	**Description**	**Example**
5	No driver intervention needed	Test vehicle adjusts following distance/brakes according to lead vehicle speed or obstacle in the driving path.
4	Driver required to input lateral control or braking to avoid collision Car performed late/hard self-braking Warning lights and sounds	Test vehicle senses obstacle and exhibits a hard brake pulse to drastically reduce speed before impact. However, the brake pulse is not enough to bring the vehicle to a complete stop, requiring the test vehicle driver to swerve or brake.
3	Driver required to input lateral control or braking to avoid collision Car performed mild self-braking Warning lights or sounds	Test vehicle senses obstacle and reduces speed but does not come to a full stop. Driver must swerve or brake to avoid collision with the soft target.
2	Warning lights or sounds	The test vehicle emits an alert that there is an obstacle in the driving path but does not take any sort of action to avoid or mitigate a collision with the soft target.
1	No car response to event	Test vehicle does not brake nor warn driver that an obstacle is in the driving path. Driver must swerve or brake hard to avoid collision with the soft target.

D. Raw Data

	Vehicle 1	Vehicle 2	Vehicle 3	Vehicle 4	Vehicle 5	Vehicle 6	Vehicle 7
ACC Curve							
Overall	2.7	2.84	2.6	2.6	2.72	2.83	2.81
Curve Degree: 101 ft.	2	2.77	2	2.08	2.18	2.33	2.25
Curve Degree: 255 ft.	3	2.92	2.67	2.83	3	3	3
Curve Degree: 230 ft.	2.75	2.83	2.75	3	3	3	3
Curve Degree: 204 ft.	2.83	2.83	3	2.5	3	3	3
Long Headway	2.64	2.72	2.46	2.5	2.67	2.75	2.79
Medium Headway	2.77	2.96	2.75	2.71	2.75	2.92	2.83
ACC Cut-In							
Overall	2.13	2.54	2.34	1.46	1.11	3	3
Cut-In Speed: 30 mph	2.25	2.5	1.5	1.75	1	3	3
Cut-In Speed: 40 mph	2.13	2.38	2.63	1.38	1.33	3	3
Cut-In Speed: 50 mph	2	2.75	3	1.25	1	3	3
Long Headway	2.25	2.42	2.42	1.5	1.17	3	3
Medium Headway	2	2.67	2.33	1.42	1	3	3
ACC Cut-Out & Reveal							
Overall	5	3.8	4.67	3.54	4.75	5	4.56
Traveling speed: 20 mph, Revealed vehicle speed: 15 mph	5	5	5	4.5	5	5	5
Traveling speed: 20 mph, Revealed vehicle speed: 10 mph	5	3.5	5	5	5	5	5
Traveling speed: 20 mph, Revealed vehicle speed: 0 mph	5	3.5	5	2.25	5	5	5
Traveling speed: 30 mph, Revealed vehicle speed: 0 mph	5	3.5	3	2	3.5	5	3.29
Traveling speed: 40 mph, Revealed vehicle speed: 15 mph	5	4.25	5	5	5	5	5
Traveling speed: 40 mph, Revealed vehicle speed: 10 mph	5	3.2	5	2.5	5	5	5
Long Headway	5	5	4.67	3.58	5	5	4.89
Medium Headway	5	2.69	4.67	3.5	4.5	5	4.23
ACC Stop & Go							
Overall	2.88	3	2.94	2	3	3	3
Long Headway	2.75	3	2.88	2	3	3	3
Medium Headway	3	3	3	2	3	3	3
AEB Obstacle							
Overall	2.7	2.17	3.3	1.73	3.53	1.13	2.63
Target: Static, Foam Car	2.33	1.33	4.67	4.33	3	1.57	1
Target: Static, Large Pedestrian	2.83	2.29	2.33	1.33	3.63	1	3
Target: Dynamic, Large Pedestrian	3	2.83	2.83	1	3.57	1	3.67
Target: Static, Small Pedestrian	3.17	2.5	3.33	1	4	1	1.83
Target: Dynamic, Small Pedestrian	2.17	2.29	3.33	1	4	1	3.67
Approach Speed: 25 mph	3.4	2.18	3.3	1.7	3.54	1.1	3.4
Approach Speed: 35 mph	3.3	2.43	3.4	1.8	3.5	1	2.5
Approach Speed: 45 mph	1.4	1.8	3.2	1.7	3.57	1.27	2
Lane Obstruction							
Overall	2.64	1	1.83	1.7	2.07	3.4	1.87
Obstacle: Barrel Cone	1	1	1	1	1	1	1
Obstacle: Foam Car	4.28	1	2.67	2.17	2.67	5	2.44
Lane Shift							
Overall	2.68	1.13	1	1.11	1	1	1
Speed: 25 mph, Shift Type: Full	2	1.08	1	1	1	1	1
Speed: 25 mph, Shift Type: Half	4.43	1.23	1	1.33	1	1	1
Speed: 35 mph, Shift Type: Full	2	1	1	1	1	1	1
Speed: 45 mph, Shift Type: Half	2	1	1	1	1	1	1
Long Headway	2.67	1.09	1	1	1	1	1
Medium Headway	2.63	1.07	1	1.17	1	1	1
Short Headway	2.75	1.07	1	1.17	1	1	1

E. Statistical Results

In the analysis of variance (ANOVA), if the value of F was greater than the value of F_{crit}, it indicated a statistically significant difference in performance scores. A t-test was then conducted on these subsets to identify which variable caused this significant difference. If the p-value (two-tailed) was less than 0.05, it indicated a statistically significant difference. All significant values are highlighted below.

ACC Curve

ANOVA: Curve Angle						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	2.37	3	0.788	22.9	3.17E-07	3.01

T-tests: Two-Sample Assuming Equal Variances

Curve Angle 1	**Curve Angle 2**	**Resulting P-value (two-tailed) from T-test**
108 ft.	301 ft.	5.15E-05
108 ft.	295 ft.	5.80E-05
108 ft.	205 ft.	0.0002
301 ft.	295 ft.	0.850
301 ft.	205 ft.	0.670
295 ft.	205 ft.	0.778

ANOVA: Headway Setting						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	0.096	1	0.096	7.83	0.016	4.75

T-tests: Two-Sample Assuming Equal Variances

Headway Setting 1	**Headway Setting 2**	**Resulting P-value (two-tailed) from T-test**
Long	Medium	0.016

ACC Cut-In

ANOVA: Event Speed						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	0.083	2	0.042	0.068	0.934	3.56

ANOVA: Headway Setting						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	0.008	1	0.008	0.015	0.904	4.75

ACC Cut-Out & Reveal

ANOVA: Revealed Vehicle Speed						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	8.62	5	1.72	2.62	0.04	2.48

T-tests: Two-Sample Assuming Equal Variances

Traveling Speed/Revealed Speed Combination 1	**Traveling Speed/Revealed Speed Combination 2**	**Resulting P-value (two-tailed) from T-test**
20 mph/15 mph	20 mph/10 mph	0.539
20 mph/15 mph	20 mph/0 mph	0.227
20 mph/15 mph	30 mph/0 mph	0.008
20 mph/15 mph	40 mph/15 mph	0.786
20 mph/15 mph	40 mph/10 mph	0.210
20 mph/10 mph	20 mph/0 mph	0.417
20 mph/10 mph	30 mph/0 mph	0.025
20 mph/10 mph	40 mph/15 mph	0.663
20 mph/0 mph	30 mph/0 mph	0.204

ANOVA: Headway Setting						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	0.900	1	0.900	1.80	0.205	4.75

ACC Stop & Go

ANOVA: Headway Setting						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	0.010	1	0.010	0.070	0.795	4.75

AEB Obstacle

ANOVA: Obstacle Type						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	0.320	4	0.080	0.056	0.994	2.69

ANOVA: Event Speed						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	1.09	2	0.547	0.621	0.548	3.56

Lane Obstruction

ANOVA: Obstacle Type						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	12.5	1	12.5	13.9	0.003	4.75

ANOVA: Headway Setting						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	0.590	2	0.295	0.144	0.867	3.55

ANOVA: Event Speed						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	1.63	2	0.817	0.389	0.684	3.55

Lane Shift

ANOVA: Shift Type						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	0.899	3	0.300	0.566	0.643	3.03

ANOVA: Headway Setting						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	0.004	2	0.002	0.005	0.995	3.55