Investigation of Smart Work Zone Technologies Using Mixed Simulator and Field Studies

---

A Dissertation

presented to

the Faculty Committee of Civil Engineering Department

at the University of Missouri-Columbia

---

by

Siyang Zhang

Carlos Sun, Ph. D., P. E., J. D., Dissertation Advisor

May 2020

APPROVAL PAGE

The undersigned, appointed by the dean of the Graduate School, have examined the dissertation entitled

Investigation of Smart Work Zone Technologies Using Mixed Simulator and Field Studies

presented by Siyang Zhang, a candidate for the degree of Doctor of Philosophy,

and hereby certify that, in their opinion, it is worthy of acceptance.

Dr. Carlos Sun

Dr. Timothy Matisziw

Dr. Praveen Edara

Prof. Henry Brown

ACKNOWLEDGEMENTS

I would like to thank my advisor Dr. Carlos Sun for his instruction and training in the past five years, he is always patient, kind, and considerate. Without him, there is no way that I could make through the PhD process. He has been and will continue to be my role model in my future career.

I would like to thank my committee member Dr. Preveen Edara for his kindness and selflessness. He is always considerate, gentle, and supportive. I would like to thank my other committee member Prof. Henry Brown, he is the one who takes care of my writings and tolerates my poor written English. I would not be able to deliver any technical reports without him. I would like to thank Dr. Matisziw for being a humor technical support. He is the one who guided me and led me into the world of civil engineering.

Special thank you goes to Zhu Qing, my previous colleague, a good friend. We have been collaborating since the first project. He always got my back.

I would like to thank Prof. Charles Nemmers for his generous and never-faded smiles, he is my role model of life. Thank Dr. Yaw Adu-Gyamfi, and Dr. Sabreena Anowar for their help and support. I also would like to thank Dr. Debra Perkowski, Dr. Isabella Zanilett, Dr. Paul Speckman, and Dr. Lori Thombs, they opened the gate to statistics to me and guide me through the pain and obstacles.

I would like to thank MoDOT and FHWA for sponsoring projects, especially Jen Harper for taking care of cases. I would like to thank everyone in ZouTrans lab, it feels like a family that always provide happiness, warmth, and support.

Finally, I would like to thank my friends and family for always being supportive. You raise me up and make me a better person.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Safety is the top concern in transportation, especially in work zones, as work zones deviate from regular driving environment and driver behavior is very different. In order to protect workers and create a safer work zone environment, new technologies are proposed by agencies and deployed to work zones, however, some are without scientific study before deployment. Therefore, quantitative studies need to be conducted to show the effectiveness of technologies. Driving simulator is a safe and cost-effective way to test effectiveness of new designs and compare different configurations. Field study is another scientific way of testing, as it provides absolute validity, while simulator study provides relative validity. The synergy of field and simulator studies construct a precise experiment as field study calibrates simulator design and validates simulator results. Two main projects, Evaluation of Automated Flagger Assistance Devices (AFADs), and Evaluation of Green Lights on Truck-Mounted Attenuator (TMA), are discussed in this dissertation to illustrate the investigation of smart work zone technologies using mixed simulator and field studies, along with one simulator project investigating interaction between human driven car and autonomous truck platoon in work zones. Both field and simulator studies indicated that AFADs improved stationary work zone safety by enhancing visibility, isolating workers from immediate traffic, and conveying clear guidance message to traffic. The results of green light on TMAs implied an inverse relationship between visibility/awareness of work zone and arrow board recognition/easy on eyes, but did not show if any of the light configurations is superior. Results anticipated for autonomous truck platoon in work zones are drivers behave more uniformly after being educated about the meaning of signage displayed on the back of truck, and performance measured with signage would be more preferable than those without signage. Applications of statistics are extension of studies, including experimental design, survey

design, and data analysis. Data obtained from AFAD and Green Light projects were utilized to illustrate the methodologies of data analysis and model building, which incorporated simulator data, biofeedback and survey response to interpret the relationship among driver perspective and mental status, and driving behavior. From the studies conducted, it could be concluded that mixed simulator and field study is a good fit for smart work zone technologies investigation. Simulators provide a safe environment, flexibility and cost-effectiveness, while field studies calibrate and validate simulator setup and its results. The collaboration of two forms of study generates legitimate and convincing results for investigations. Applying statistical methodologies into transportation simulator and field studies is a good way to make experiment and survey design more rational, and the statistical methods are applicable for further data analysis.

CHAPTER 1: INTRODUCTION

1.1 OVERVIEW

Work zone safety is a critical concern in transportation, because driving behavior would be very different as work zone deviates from regular driving environment. Distracted driving is playing a significant role in crashes. According to National Highway Traffic Safety Administration (NHSTA 2019), distraction counted for 9% of fatal crashes, with 3,166 fatality due to distracted driving in United States in 2017. According to Missouri Department of Transportation, 139 Truck-Mounted Attenuator (TMA) crashes occurred from 2012 to 2017, with 43.9% caused by distracted driving.

The goal of this dissertation study is to improve work zone safety, and the main objective is to investigate smart work zones using mixed simulator and field studies. Original and significant contributions of this study include: 1) investigation of smart work zone technologies to improve safety & efficiency, 2) relationship between simulator and field studies and their synergy, 3) design of simulator studies for smart work zone investigations, 4) utilization of psychophysics and biofeedback in transportation, 5) giant model incorporating quantitative and qualitative performance measures.

1.1.1 Investigation of Smart Work Zone Technologies

Three types of new technologies were investigated for work zones: Automated Flagger Assistance Devices (AFADs), green lights on Truck-Mounted Attenuators, and autonomous truck platoons. AFADs are deployed in stationary work zones, green light TMAs are utilized during mobile operations, and autonomous truck platoons communicate with human driven vehicles in work zones.

Generally speaking, AFADs and autonomous truck platoons are more suitable than green lights on TMAs to be tested in a transportation simulator, as both of them are testing for relative macroscopic factors, and relative validity provides useful information. Testing light configuration in simulator is a challenge, as the presence of light would be different from reality due to various reasons, for example light intensity, light color, or resolution and tone of monitor.

1.1.2 Relationship between Simulator and Field Studies

Using different orders of conducting simulator study and field study may affect the whole research project. Field studies could help calibrate and validate simulator study, and simulator results indicate clearer testing direction of field study.

Aspects of simulator calibration includes size, position and tone of signage, speed of acceleration and deceleration, steering wheel control, etc. Conducting field study first would provide realistic information and methodology to develop simulator scenarios. Also, for proposed configurations that are not being tested in the field, they can be programmed in simulators.

Forming field studies from simulator results has the advantage of being cost-effective, as the results narrow down configurations to be further evaluated. Although the simulator results are relative validities, the field studies are validating the results in reality and provide absolute validity. Collaborating simulator and field studies generates more solid and legitimate results.

1.1.3 Design of Simulator Studies for Smart Work Zone Investigations

Design of work zone elements adjust to projects specifically. In AFAD project, special elements include the MoDOT hybrid AFAD with and without Changeable Message Sign (CMS) and the content on CMS, two-way two-lane rural highway and a temporary work zone with one lane closure, human-animated flagger with gesture and STOP SLOW sign. In the green light TMAs project, both daytime and nighttime scenarios are coded, with different light configurations flashing at specific frequency, and work vehicles moving continuously as mobile work zones. A water bottle rolling with the work vehicle is essential as well to test for disability glare. Autonomous truck platoon in work zone is more complicated, as the truck simulator is federated with a car simulator via virtual network wirelessly. Exact scenarios are undetermined at this moment.

Besides scenario development, there are other factors to be considered, for example fidelity of simulator and scenes, potential simulator sickness, an appropriate length of time to avoid sickness, and order of scenarios and sequence bias. Human factors related issues need to be considered as well.

1.1.4 Psychophysics and Biofeedback Utilization

In simulator studies, biofeedback devices can be used utilized to measure drivers' stress level when driving though different designs and configurations. Eye trackers were utilized in three projects related to driving simulators discussed in this dissertation, and a wristband measuring heart rate, blood pressure and stress level is going to be used in the autonomous truck platoon simulator project. By combining driver behavior and driver's real-time stress level, researchers could better understand the relationship between driver's mental status and physical movement. Biofeedback data helps further interpret the design of signage, impact of configurations, and effectiveness of education.

1.1.5 Comprehensive Model Incorporating Quantitative and Qualitative Performance Measures

Survey responses are subjective, and they are a qualitative performance measure instead of quantitative data. Thus, the statistical technique applied to data analysis of survey responses is different from field and simulator data. Commonly used methods include nonparametric Mann-

Whitney test for ordinal data (De Winter & Dodou 2010), and Friedman test for ranking preference (Friedman 1940).

Currently, a large part of studies in transportation divide quantitative data and qualitative data collected and run analysis separately. Finding ways to integrate subjective data and objective survey response into a giant model would be very helpful. For example, by tying a simulator participant's survey response to his/her trial performance, researchers could see a clearer relationship between driver's perspective of design and driving behavior, which helps interpret the relationship between driver performance and public acceptance on certain new configurations or regulations.

1.2 INVESTIGATION OF SMART WORK ZONE TECHNOLOGIES TO IMPROVE DAFETY & EFFICIENCY

A work zone is a construction area of roadway. It is common that during construction period, shoulder or lanes are closed, which slows down traffic flow and causes congestion, even crashes. Enhancing safety and efficiency of work zones is of high importance. According to FHWA (2009), there are two types of traditional work zones: stationary and mobile.

1.2.1 Automated Flagger Assistance Devices (AFADs) for Stationary Work zones

A stationary work zone occupies a certain location during construction. Traffic engineers have proposed different methods to slow down the approaching speed and extend the merge distance of vehicles as they approach stationary work zones. Studies investigating new technology include alternative merge signs (Zhu et al. 2015), automated traffic light systems (Subramaniam et al. 2010), flashing STOP/SLOW paddles (Pigman et al. 2006), Remote Controlled (RC) Flagman (Jessberger 1999), IntelliStrobe Safety Systems (Missouri Department of Transportation 2006) and other types of AFADs (Cottrell Jr 2006; Finley et al. 2011; Terhaar 2014).

Flaggers are professionally trained to guide and direct vehicles through stationary work zones, but they are often located closest to the oncoming traffic. As a result, they are exposed to risks associated with errant drivers (Antonucci et al. 2005). Studies have shown that a very high percentage of work-related crashes occurred in the advance warning area where flaggers are located (Ishak et al. 2012; Srinivasan et al. 2007). Therefore, discovering ways to protect flaggers is an important issue in work zone safety. To reduce exposure to traffic and improve flagger safety, there are several countermeasures applied in work zones, including the use of buffer spaces and barriers (Trout and Ullman 1997). One countermeasure that can be applied when there is a one lane closure on a two-lane road is an Automated Flagger Assistance Device (AFAD). An AFAD removes a worker from having to be near the approaching traffic at a work zone. MoDOT proposed a new design of AFAD by having STOP/SLOW paddles with supplemental flashing Red/Yellow lights, with a changeable message sign attached, and the whole AFAD is mounted on TMA. A project was conducted to evaluate the configurations. The project contained two phrases, with one field study and one simulator study.

1.2.2 Green Light Truck-Mounted Attenuators (TMAs) for Mobile Work Zones

Mobile work zones for various types of moving operations such as striping, sweeping, and pothole filling are an important component of maintaining highways. The Manual on Uniform Traffic Control Devices (MUTCD) (FHWA 2009) provides guidance for the layout for mobile work zones, using shadow vehicles, arrow boards, and signs to warn drivers that they are approaching a mobile work zone. Research of mobile work zone safety focuses drivers' attention catching, early recognition of work zone, and notification of short distance between work zone and approaching vehicles. Studies on new devices such as mobile work zone alarm systems (Brown et al. 2015) were conducted previously. In addition, a Truck-Mounted Attenuator (TMA) attached to a construction vehicle helps to mitigate the impact of a collision from a highway vehicle that fails to recognize the mobile work zone. Amber/white lights are typically used on the TMA to help draw motorists' attention to the moving work zone. Despite these precautions, some drivers do not respond to warnings and collide with the TMA. Distracted driving may be a contributing factor in these collisions.

Research was conducted by the Virginia Transportation Research Council (VTRC) on TMA crashes in work zones in Virginia (Cottrell Jr 2015). The goals of the VTRC research were to find trends in TMA crashes over a period of three to five years and find out the biggest causes of TMA crashes in work zones. From 2011-2014, the number of TMA crashes had increased in contractor work zones, and the Virginia DOT work zones experienced approximately the same level of TMA crashes per year. The study found that some of the leading contributing factors to TMA crashes in work zones were distracted driving, sight distance issues, and improper placement of the TMA in work zones.

TMA Incidents

There were 139 TMA crashes reported in Missouri from 2012 to 2017. Thus, there was a TMA hit approximately every 15 days. The major reason for drivers to hit a TMA was distracted driving (44 percent), as shown in Table 1.2.2.1. Among all TMA crashes, 97 were related to mobile work zones (70 percent), 20 were stationary work zones, and the others unknown. Some TMA operations had a higher risk of crash, such as pothole patching, striping, and sweeping, as shown in Table 1.2.2.2. This result could be due to higher exposure, or because these specific operations were riskier. The data show that vehicles crashing into a TMA at a mobile work zone is a real concern.

Table 1.2.2.1 Reasons for TMA crashes in Missouri

	Distracted driving	Late merging	Speeding	Others	Not reported
Count	61	21	7	7	43
Percentage	43.9%	15.1%	5.0%	5.0%	30.9%

Table 1.2.2.2 Mobile work activity when TMAs were struck in Missouri

Operation	Count	Percentage
Pothole patching	31	22.3%
Striping	20	14.4%
Sweeping	18	12.9%
Maintenance (not specified)	13	9.4%
Bridge	9	6.5%
Mowing	8	5.8%
Cleaning dirt	8	5.8%
Signage	7	5.0%
Spraying weeds	3	2.2%
Rolling	2	1.4%
Other	9	6.5%
Unknown	11	7.9%

The aforementioned data shows that vehicles crashing into a TMA at a mobile work zone is a real concern. MoDOT was motivated to conduct a green lights on TMAs project to reduce TMA crashes and improve mobile work zone safety. By testing different combinations of light colors, the objective of this project was to improve mobile work zone safety and evaluate the performance of green lights on TMAs. This project consisted of two phases: Phase One involved simulator testing, with four different configurations followed by post-simulator surveys; Phase Two involved a field study to compare the performance of green only and amber/white color lights on TMAs.

1.2.3 Autonomous Truck Platoon in Work Zones

Autonomous Truck Platooning with low level automation is expected to be widely deployed in the next five years. Potential benefits include fuel consumption reduction, higher efficiency due to shorter headway and more relaxed truck drivers, and lower freight cost. However, public acceptance of autonomous truck platoon is a big challenge, because of novelty effect and higher fatal rate of crash especially for crashes involving heavy trucks in work zones (FMCSA 2014), as shown in Table 1.2.3.1. Previous simulator research shows that external displays used to enhance communications between autonomous vehicles and pedestrians or cyclists can be adapted to truck platooning, and special education of truck platoon signage is helpful to guide human drivers (Schoelz et al. 2019).

Table 1.2.3.1 Fatal Crashes Involving Large Trucks in US in 2012

Crash Type	All Fatal Crashes	Work Zone Fatal Crashes
Involved at Least One Large Truck	11.20%	23.60%
Involved a Large Truck and Two or More Vehicles	16.90%	32.60%
Involved a Large Truck That Was Parked/Working	4.10%	18.90%

The statistics of trucks involved in the work zone crash fatal rate shows that safety of trucks and truck platoons is a serious issue. Studying interaction between passenger car drivers and truck drivers and enhancing their communication are of great importance. To examine how human drivers interact with autonomous truck platoons, the Smart Work Zone Deployment Initiative (SWZDI) funded a project. This project is conducted in simulator, providing safe and efficient environment to collect quantitative data and qualitative post-simulator survey response.

1.3 LIST OF PUBLICATIONS, PRESENTATIONS, AND ACADEMIC AWARDS

1.3.1 Publications

Zhang, S., Qing, Z., Brown, H., Sun, C., and Edara, P. (2019). “Simulator and field study of green lights on truck-mounted attenuators in Missouri during mobile operations.” *Transportation research record, 2673*(2), 769-778.
(Awarded as the Committee’s Best Paper of 2019 by the TRB Work Zone Traffic Control Devices Committee (AHB55))

Qing, Z., Zhang, S., Brown, H., and Sun, C. (2019). "Evaluation of Truck-Mounted Automated Flagger Assistance Devices in Missouri: Case Study." *Journal of Transportation Engineering, Part A: Systems*, *145*(12), 05019006.

(Awarded as Editor's Choice, and featured in the Editor's Choice section of the Journal of Transportation Engineering Part A: Systems)

1.3.2 Selected Presentations

Lectern, *2020 TRB Annual Meeting*, Investigation of Smart Work Zone Technologies Using Mixed Simulator and Field Studies.

Lectern, *2019 TRB Annual Meeting*, Simulator and field study of green lights on truck-mounted attenuators in Missouri during mobile operations.

Poster, *2018 TRB Annual Meeting,* Simulator and field study of truck-mounted Automated Flagger Assistance Devices in Missouri.

Poster, *ITS Heartland 2019 Annual Meeting,* Investigation of autonomous/connected vehicles in work zones.

Poster, *2018 Annual TEAM Transportation Fair,* Evaluation of green lights on truck-mounted attenuators.

Poster, *Missouri's 2018 Highway Safety and Traffic Blueprint Conference,* Evaluation of green lights on truck-mounted attenuators.

Poster, *Missouri's 2017 Highway Safety and Traffic Blueprint Conference,*Evaluation of Automated Flagger Assistance Devices.

1.3.3 Academic Awards

Helene M. Overly Memorial Scholarship, 2019

3rd place, ITS Heartland Student Poster Competition, 2019

2nd place, TEAM STL Poster Contest, 2018

1st place, Missouri's 2018 Highway Safety and Traffic Blueprint Conference Poster Competition

1st place, Missouri's 2017 Highway Safety and Traffic Blueprint Conference Poster Competition

1.4 TEACHING AND SUGGESTIONS ON CURRICULUM

The author of this dissertation used to be a graduate teaching assistant (TA) during Fall 2019 semester for class CE 3100 Fundamental of Transportation Engineering. She assisted head TA with lab sessions, helped students with GPS surveying project using AutoCAD, guided students through project of modeling traffic and signals at intersections in VISSIM, and provided instructions to students on signal timing optimization via PASSER.

As one of the students that assisted with building up transportation simulators and expanded them into a federated system with six modes, she contributed to hardware setup and software development. She observed and was involved in the incremental improvement the simulator methodology over multiple projects, and mentored graduate and undergraduate research assistants and guided them through projects.

From her own experience, she would suggest that MU Civil Engineering Department to adjust the curriculum, and introduce courses related to transportation simulator to electives. Currently, no course on Unity (the engine of simulator scenario development) is offered by the department, and students have to seek for online resources to learn Unity, which is of low efficiency and sometimes extends the project period. She strongly recommends that the department offers a course on Unity, or collaborate with the Computer Science Department to offer this course, so that students working on transportation simulator projects could be trained systematically and gain more legitimate results out of simulator studies. Besides Unity, more precise course on AutoCAD and Sketch Up are needed. Often times, sponsors of simulator studies aim to test for new configurations or designs, which are not pre-models in Unity, and need to be built by project researchers. AutoCAD enables the possibility of building up new models monochromatically while Sketch Up offers the options of painting, and the final products are able to be imported into Unity and adopted to scenarios.

Data analysis is another important part in transportation engineering. She highly recommends that students register for Statistics courses. Use courses offered by MU Statistics department as examples: STAT 7710 Intro to mathematical statistics, STAT 7540 Experimental design, STAT 7510 Applied STAT model 1 (R), STAT 8220 Applied STAT model 2 (SAS), STAT 8370 Statistical consulting. These courses would help students to better understand how to design experiment and survey, and how to apply analysis methods and build models based on different types of data.

Transportation students may be interested in Geographic Information Systems as well. These related courses are offered by Geography Department, and some of them are cross-listed with Civil Engineering Department, and cross-listed for undergraduate and graduate levels. Some recommended courses are GIS 1, GIS2, location analysis, transportation geography, and spatial analysis. If students are interested in higher level courses, then geospatial science in national security and Geoinformatics seminar are suggested, as well as remote sensing series.

1.5 SERVICE

During the years pursuing PhD, the author of this dissertation provided service to ZouTrans Lab and College of Engineering. Her major service included officer of Institute of Transportation Engineers Student Chapter in University of Missouri (MizzouITE) and officer of Central Missouri Institute of Transportation Engineers (CMITE) professional chapter.

1.5.1 Service to MizzouITE Student Chapter

The author served as a student member then as an officer in Institute of Transportation Engineers Student Chapter in University of Missouri (MizzouITE) since 2018. MizzouITE committed to serving Mizzou campus and communities, and were engaged in engineering education outreach programs to create awareness among K-12 and college students about careers in the engineering profession.

During her service as secretary in 2018 and as president in 2019 and 2020, she organized activities such as tailgates and get-together parties. She successful raised $8,840 of travel fund and supported student members' trip to Washington D.C. for Transportation Research Board (TRB) conference. Her raised fund was able to support six members' travel in 2019 and another 11 members' trip in 2020. She was responsible for writing chapter annual reports and help the organization won cash awards three times (as shown in Vita).

She hosted multiple lab tours for the department, college, campus, community and some other visits. Table 1.5.1.1 lists some of the lab tours she hosted.

Table 1.5.1.1 Lab Tours Hosted

Event	Date
Engineering Week Exhibit	3/14/2018
CMITE Chapter Visit	3/30/2018
Show Me Mizzou Open House	4/13/2018
Alumni Visit	4/27/2018
Chinese Delegation Visit	5/9/2018
ITS Visit	5/25/2018
Mizzou International Experience Day	7/24/2018
High School Ambassadors Visit	9/7/2018
Buckle Up Phone Down Day	10/19/2018
MU Extension Visit	10/24/2018
Engineering Week Exhibit	3/14/2019
Faculty Candidate Visit	4/3/2019
CMITE Chapter Visit	4/12/2019
Alumni Visit	4/27/2019
Mizzou International Experience Day	7/24/2019
Emergency Response Personnel Visit	9/16/2019
NTSB Roundtable Group Visit	10/29/2019
Faculty Candidate Visit	2/17/2020
Faculty from Tongji Simulator Lab Visit	3/8/2020

1.5.2 Service to CMITE Professional Chapter

The author started her service to CMITE since February 2018 as secretary/treasurer, stepped up as vice-president in December 2018, and became the president in December 2019. During her three-year service, she attended and hosted executive meetings, planned, announced, attended and hosted member meetings, collected dues and charges, invited speakers, arranged lunch and social activities, and prepared annual reports. CMITE chapter is going to transit into CMITE section due to ONE ITE move, and she hosted executive meetings and adjusted CMITE section bylaws.

Her dual roles of officers in MizzouITE and CMITE connected student chapter and professional chapter together tightly, making stronger connections between students and professionals in transportation, and helped students to expand their network. After PhD graduation, she would not be the president of MizzouITE student chapter anymore, but would continue to serve CMITE as president till the end of duty period.

1.5.3 Other Services

Other services provided including lab management, asset and equipment arrangement, conference trip planning, reimbursement, etc. She also mentored high school research assistant, undergraduate and graduate research assistants, and help them with research, course registration and well-beings.

CHAPTER 2: LITERATURE REVIEW

2.1 AUTOMATED FLAGGER ASSISTANCE DEVICES (AFADS)

AFADs are designed to protect flaggers in work zones by allowing flaggers to control traffic signals remotely instead of standing right next to occupied lanes. According to the *Manual on Uniform Traffic Control Devices* (FHWA 2009), there are two different types of AFADs: STOP/SLOW and red/yellow lens. Both types of AFADs are remotely controlled. In their 2005 technical provision, FHWA (2005a) regulated the maximum distance between two AFAD devices to 1,000 ft. AFADs should not be used for long term work or as regular traffic control signals. According to the FHWA (FHWA 2009), a STOP/SLOW AFAD shall include a sign with STOP and SLOW faces showing alternatively, which could be controlled remotely. A red/yellow lens AFAD shall include a steady circular red lens and a flashing circular yellow lens. A gate arm is required for the red/yellow lens AFAD, which lowers the arm to stop approaching traffic while the red lens is illuminating and raises the arm to release stopped traffic while the yellow lens is illuminating.

Some commercial STOP/SLOW AFADs include the AutoFlagger 76 (AF-76) (Figure 2.0.1.1) (Safety Technologies 2015a), and J4 Flagger Workstations. Commercial red/yellow lens AFADs include the AutoFlagger 54 (AF-54) (Figure 2.0.1.2) (Safety Technologies 2015b), RC Flagman RCF 2.4 (Figure 2.0.1.3) (North America Traffic 2016), Automated Flagger AF-100 (Synergy Technology 2017), and IntelliStrobe W1-AG (Figure 2.0.1.4) (IntelliStrobe 2017). FHWA also created a policy memorandum (FHWA 2005b) and provided technical provisions (FHWA 2005a) for AFADs. Based on the work and materials from FHWA, American Traffic Safety Services Association (ATSSA 2012) published a guidance document on AFAD usage in 2012.

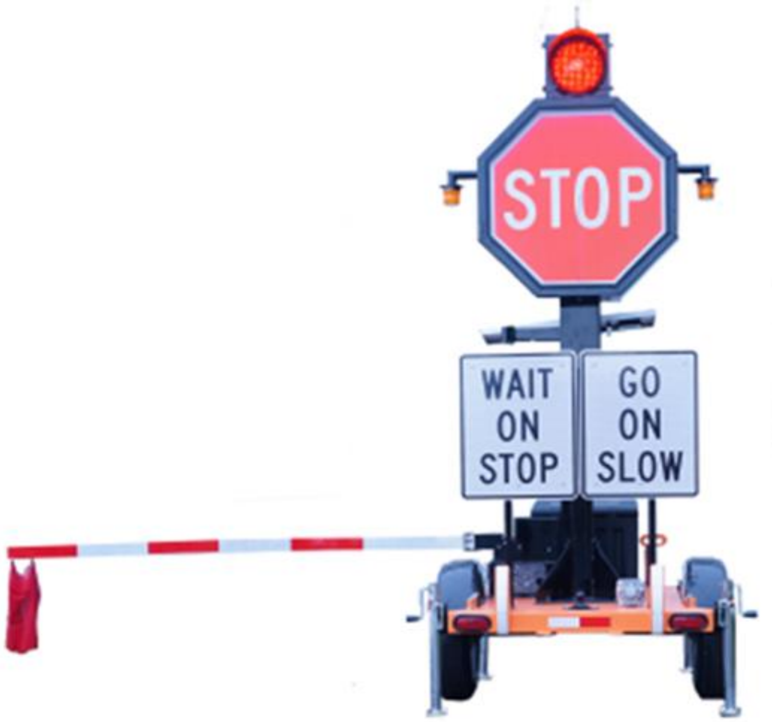


Figure 2.0.1.1 AutoFlagger 76 (Safety Technologies 2015a)

Figure 2.0.1.2 AutoFlagger 54 (Safety Technologies 2015b)

Figure 2.0.1.3 RC Flagman RCF 2.4 (North America Traffic 2016)

Figure 2.0.1.4 IntelliStrobe W1-AG (IntelliStrobe 2017)

To evaluate the effectiveness of AFAD, research and field studies were performed by the Ohio Department of Transportation (ODOT) (Jessberger 1999), Washington County (Kansas) Public Works (Harris 2002), MoDOT (Missouri Department of Transportation 2006), Minnesota Department of Transportation (MnDOT) (MnDOT 2005; Terhaar 2014), Virginia Transportation Research Council (VTRC) (Cottrell Jr 2006), and Texas A&M Transportation Institute (TTI) (Finley 2013; Finley et al. 2011; Trout et al. 2013). Some of these evaluations are discussed in the following sections.

2.1.1 STOP/SLOW AFADs

MnDOT (2005) tested the AutoFlagger traffic control devices in late 1990s as an enhancement to flagging systems. A human operator controlled the AutoFlagger devices in both directions remotely. Surveys were gathered from drivers and operators on their opinions of the AutoFlagger and the responses were positive.

VTRC and the Virginia Department of Transportation (VDOT) reviewed applications of AFADs in Minnesota (MnDOT 2005) and evaluated AutoFlagger deployments in two areas. VTRC compared AutoFlagger with other AFAD systems (Cottrell Jr 2006). The AutoFlagger deployed in Virginia was a STOP/SLOW paddle device equipped with a horn for warning purpose. The first deployment was located in the Wytheville area. The AutoFlagger was deployed under different types of construction and maintenance projects and was also displayed at a safety day. VDOT used it on roadways with narrow shoulders or no shoulders by putting the device in the lane with a 50-ft taper of cones in front of it. The deployments in Wytheville showed that the "WAIT ON STOP – GO ON SLOW" signs were misunderstood by drivers due to the novelty of AutoFlagger. The second deployment was located in the Beach area. In contrast to the deployment in Wytheville, staff in Beach felt more comfortable using AutoFlagger in long straight areas with wide shoulders and clear sight distance, rather than in areas with narrow shoulders. The crews also suggested that horns should be made louder in order to be heard, and flashing lights should be larger and brighter to enhance visibility. VTRC concluded that although the deployment of AutoFlagger is limited by shoulder conditions, the application of AutoFlagger provides a safe work zone environment, costs less labor, and saves money for the long term. A drawback is that it may be harder for drivers to locate a flagger in order to communicate with him/her.

2.1.2 Red/Yellow Lens AFADs

In the late 1990s, the Ohio Department of Transportation (ODOT) (Jessberger 1999) evaluated the Remote Controlled (RC) Flagman. The RC Flagman device was placed at a two-lane highway location, with one lane closed, and the evaluation involved comments from ODOT employees who operated the devices, public interviews, an assessment of cost effectiveness, and accident statistics. The RC Flagman contains red/yellow signals mounted on a mobile trailer, a remote control unit, a gate arm, and a "STOP HERE ON RED SIGNAL" paddle. During the ODOT field experiment, operators had trouble with short battery life and weak button contact of the remote control units, time delay from the pressing of the button to the changing of the signal light, and slow movement of the gate arm motors. Operators also recommended that the visibility of gate arm be enhanced. Operators indicated that the set up and operation of RC flagman was easy, and they were satisfied with driver reactions. Most of the motorists thought that the device was visible, the STOP message was presented clearly, and the RC Flagman freed one flagger and provided a safer environment for flaggers. Some interviewees were worried about flaggers losing jobs. Although some problems existed, thee overall comments from ODOT operators and public regarding the use of RC Flagman were favorable. Based on cost and maintenance history, the study found that using RC Flagman is cost effective. No accidents were reported during the two-year evaluation of RC Flagman. Thus, ODOT concluded that RC Flagman is at least as safe as traditional flaggers. Similar to Ohio, the RC Flagman evaluation in Washington County, Kansas, also found that it is cost-effective, and the visibility of a red light makes it work even better than human flaggers (Harris 2002). According to RC Flagman, no accidents have been reported at RC Flagman sites since this device was produced in 1993 (Harris 2002).

In Fall 2005, MoDOT piloted the IntelliStrobe flagging system (IntelliStrobe 2017) in the South Central District. An IntelliStrobe device contains red/yellow signals, two remote control units to be used by one person, and a danger alert. The yellow light flashes continually, and when the red light turns on, the gate arm lowers to stop approaching traffic. In case motorists misunderstand or violate the signal and "do not stop" sign, the danger alert sounds to alert operators. The IntelliStrobe Safety System is suitable for short work zones since one flagger controls both ends. It frees up one flagger, and as a result, shortens the time needed for deployment (Missouri Department of Transportation 2006).

2.1.3 Both STOP/SLOW and Red/Yellow Lens

In addition to the evaluation of STOP/SLOW AFADs performed in 2005, MnDOT (Terhaar 2014) held two training sessions for its employees in 2013 to further investigate and evaluate AFADs. These sessions include introduction and demonstration of AFADs, set up, operation, and take down, discussion of impressions and limitations, and field tests. Both AutoFlagger AF-76 (STOP/SLOW) and AutoFlagger AF-54 (red/yellow lens) were evaluated. The outcome of this study indicated that a set of AFADs could be operated by one or two personnel remotely from traffic, and maintenance staff were willing to use AFAD overall. Setting up and taking down AFAD requires more time and effort than traditional flagging. The result also suggested that AF-76 fits in wide shoulder work zones, while AF-54 fits in narrow shoulder locations, and both AF-76 and AF-54 are recommended for two-lane highways.

The review of the previous AFAD evaluations showed that there was very little use of quantitative performance measures and no applications of statistical methods. The lack of previous scientific AFAD evaluations is a major motivation for the present MoDOT study which uses quantitative performance measures such as speeds, stop locations, wait times, reaction times, and noncompliance rates in addition to surveys. Another motivation of this study is that MoDOT proposed a new design of AFAD by adding CMS to the AFAD, and mounting the AFAD on TMA. The present study also uses statistical techniques for analyzing and interpreting quantitative and qualitative performance measures.

2.2 LIGHT CONFIGURATIONS

This review of the existing literature on the use of green lights on construction vehicles includes aspects of light colors, light positions, and other light factors which are described in the following sections. Light color is one important factor in TMA light bar configuration, as the sensitivity of human vision varies across colors. Several research studies on service vehicle warning lights have been conducted. One such study was performed by the American Association of State and Highway Transportation Officials (AASHTO 2009). AASHTO investigated and defined the best practices for selecting warning lights on roadway operations equipment. They considered safety and lighting issues, along with defining lighting selection. The study found that an asynchronous flashing pattern was the most effective. Amber/white colored lights were also proven to be more effective than blue and red lights. There was no observed difference between different types of light sources, such as LED or strobe. The study also recommended that lights be placed with a portion of the vehicle behind them. Finally, lights with a higher effective intensity obtained better drivers' attention. However, safety and time of day should also be considered. The glare of high-intensity lights may affect driver vision. This effect can be mitigated by adjusting the level of lighting and by establishing a difference between daytime and nighttime lighting configurations.

A National Cooperative Highway Research Program (NCHRP) study was conducted to evaluate the effectiveness of warning lights on roadway maintenance vehicles and their impacts on motorist awareness (Gibbons 2008). Photometric characterization, static screening, and performance experiments were used in order to review current practices and further investigate different warning light systems. The study determined that drivers noticed flashing lights more quickly than steady lights. Furthermore, flashing lights with an asynchronous pattern were proven to be more effective than flashing lights with a synchronous pattern. An asynchronous flashing pattern using amber and white lights was found to be the most effective, as it was better recognized by drivers than red and blue ones. In addition to color, the type of light made an impact on obtaining the attention of the driver. The study found that, for halogen and strobe lights, increasing the intensity helped drivers on the road to recognize them much more readily. In addition, lane change distances for LED warning lights were much farther than lane change distances for warning beacons mounted high or low on highway maintenance vehicles. The report recommended that the warning light system should also have a higher effective intensity in the daytime, compared to nighttime, and the light should be laid out on a controlled background. It was additionally determined that lights with a higher effective intensity are necessary in adverse weather conditions but result in increased glare. Overall, lighting characteristics and layout along with environmental conditions are all factors to consider while designing a lighting system. Furthermore, it is important to evaluate the measurement techniques used for lighting systems and to standardize them across different manufacturers. This report pointed out that during daytime, green light had the shortest detection time, but it did not perform well with respect to disability glare and discomfort glare. The balance between conspicuity and distraction of warning beacons needed to be maintained, as well as the balance between high effective light intensity and disability glare issues. The report recommended placing light bars at a higher position to reduce the glare issue, and recommended the amber/white combination for maintenance vehicles to avoid confusion.

Internationally, a research study was performed in the United Kingdom regarding warning beacons (Cook et al. 2000). Researchers investigated the conspicuity of warning beacons with lights consisting of different flash types, flash rates, and flash intensities. Through laboratory and field trials, the study found that strobe lights yielded a greater sense of urgency for drivers. Also, flash rates of 4 Hz were found to be ideal in getting driver's attention. Finally, high flash intensities were determined to minimize detection time of the warning beacon.

2.2.1 DOT Practices

The Texas Department of Transportation (TxDOT) (Ullman and Lewis 1998) performed research evaluating the vehicle fleet warning lights in order to determine if using different colored warning lights on fleet vehicles, other than the standard flashing yellow lights, provided greater safety in highway work zones. This study found out that 12 states used colors other than yellow-only on some equipment. The combinations of colored lights tested were yellow/blue and yellow/blue/red. The results showed that drivers perceived flashing yellow lights to convey a less hazardous situation, which may not be the correct hazardous level, compared with flashing yellow/blue or yellow/blue/red combinations.

The Kentucky Transportation Center (KTC) surveyed state DOTs on warning light color options for work vehicles and their effectiveness (Howell et al. 2015) and received 16 responses. The survey indicated that four states used blue lights on their maintenance vehicles and five different states used red lights on their work vehicles. The study reviewed warning light practices at other state DOTs nationwide as well as other relevant factors, such as light source, light color, and layout of lights. This review served to gather information on the current state of affairs on work vehicle warning lights and allow for Kentucky Transportation Cabinet (KYTC) to improve their work vehicle lighting systems. All state DOTs surveyed had policies and regulations covering their work vehicle lighting programs. They all reported using amber colored lights and LED lights on their work vehicles. Other commonly used light colors were white, and red and blue for emergency vehicles. The DOTs also reported using different lighting intensities at day and night, along with varying light colors by highway vehicle type. A strong preference for roof-top mounted lighting was also indicated. The summary of light usage for selected states is shown in Table 2.2.1.1. This research found that most of the agencies that participated in the survey strongly preferred putting warning lights on top of the roof of highway maintenance vehicles. The study also found that when drivers encountered red and blue warning lights on highway work vehicles, they linked the red and blue lights with other emergency vehicles. The KYTC plans to use this knowledge to place appropriate lighting on their emergency and public safety vehicles. KYTC attempted to use amber and green combination, however they were not available from manufacturers.

Table 2.2.1.1 Color usage on emergency and warning vehicles by state

State	Maintenance	Emergency vehicles	Note
Alaska	amber, blue	n/a	use different colors for maintenance vs. emergency
Illinois	n/a	amber	prevent glare
Indiana	amber	n/a	tested range of weather and lighting conditions
Iowa	amber	n/a	can use white, blue, red to complement amber
Maine	amber	n/a	amber for maintenance
Massachusetts	amber, white	red	use different colors for maintenance vs. emergency
Michigan	amber	n/a	amber for maintenance
Minnesota	amber, blue	n/a	to promote safety
Missouri	amber, white	red, blue	use different colors for maintenance vs. emergency
New Hampshire	amber	n/a	amber for maintenance
Ohio	green, white, amber	n/a	to improve truck visibility
Oklahoma	amber, blue, red, white	n/a	4 colors used for maintenance
South Dakota	amber	n/a	amber for maintenance
Texas	Amber/yellow	red	use different colors for maintenance vs. emergency
Washington	amber, red	blue	use different colors for maintenance vs. emergency

There are existing laws in place in each state that restrict the color and type of light that can be used on emergency, road maintenance, and other vehicles. In general, red and blue warning lights are always allowed for emergency vehicles, and a limited number of red/blue lights can sometimes be used for other service vehicles like highway maintenance vehicles or snow trucks. All states allow amber colored lights on maintenance vehicles.

The Ohio Department of Transportation (ODOT) is the first state to implement green lights on work vehicles. The ODOT use of green lights is limited by statute to snow removal trucks (ODOT 2013), as shown in Figure 2.2.1.1. The color combination that the ODOT implemented was amber, green, and white. Under the configuration, three ambers are always lit and alternate with three whites or three greens. In a phone interview with the researchers, ODOT personnel indicated that this configuration was determined partly through a survey of trucks set up with different colors at the state fair. ODOT found that a high flash rate was necessary to prevent colors from blending together and sickness from looking at the lights. The deployment found green lights to be more effective in catching the driver's attention. ODOT predicted that the green lights would help make roads safer because drivers would be able to easily spot work vehicles in snowy weather. ODOT is still in the processing of evaluating the safety impacts of using the green lights but believes that they help to reduce crashes.

Figure 2.2.1.1 ODOT green light snow-removal truck (ODOT 2013)

Personnel from the Michigan Department of Transportation (MDOT) indicated in a phone interview with the researchers that MDOT is in the process of implementing amber and green lights on its snow removal vehicles. Details regarding the light configuration are still being finalized. It is expected that this implementation will be completed in fall 2018. The Michigan statute passed allows for MDOT to use green lights on both winter maintenance vehicles and regular maintenance vehicles, but MDOT is currently limiting its use to snow removal vehicles.

The review of previous research and DOT practices show that light bar configurations on TMAs were not evaluated, and no quantitative measure or statistical analysis was done to examine them. With the lack of evaluations of scientific light configurations, MoDOT was motivated to conduct a study of the green lights on TMAs using quantitative performance measures such as speeds, merge distance, work zone and arrow direction recognition distance, and disability glare along with surveys. Statistical techniques are also applied in the study to analyze and interpret quantitative and qualitative performance measures.

2.3 TRUCK PLATOON IN WORK ZONES

As large trucks involved crashes have higher fatal rate than regular crashes, Osman (2016) built a generalized ordered response logit (GORL) model to analyze the injury severity of large truck crashes in work zones, finding out that daytime crashes, no control of access, higher speed limit, and rural principal arterials were the significant factors to severe injuries. However, most work zone safety analyses and improvements focus on passenger vehicles and treat trucks and buses as multiple passenger cars by considering only the size of the vehicles, which is not appropriate. In order to improve truck safety in work zones, FHWA (2016) published guidance specifically for trucks. FHWA proposed changes to work zone design, for example adding trucks to traffic analysis, maintaining sufficient lane width, signage, and rest area, splitting truck-only lane through work zone, and enforcing work zone speed. FHWA also suggested ways to help drivers understand signage and negotiate with surroundings, including encouraging safe driving behavior, adopting work zone ITS technology and advanced traffic control devices.

With the current Connected and Autonomous Vehicles, enhancing work zone detection is one way to reduce risk and improve safety. Lee et al. (2013) proposed a Kernel-based traffic sign tracking system to help improving work zone recognition by detecting and analyzing the shape and color of signage and generating combined kernel. Results showed that this method picked up missed signs which were missed by detectors and improved sign detection and classification performance. Seo et al. (2015) conducted further research based on Kernel-based tracking system and added computer vision methods to recognize and identify work zone boundary and signs, and the method was validated by analyzing video data under various weather conditions.

Communication in work zones is as important as detection. Southwest Research Institute (SwRI) and Texas Department of Transportation initiated a program on Cooperative Autonomous Vehicle Systems in 2006 (Brown et al. 2014), developing unmanned ground vehicle technology, enhancing work zone worker gesture communication, and adding local infrastructure-based devices and SwRI cameras to enhance local communication and improve work zone safety.

Besides traffic and drivers' safety, the safety of work crew is of high importance. In a Strategic Highway Research Program (SHRP) report, it was suggested that worker exposure to traffic should be reduced, and an automated follow vehicle system including mobile opto-electronic beacon, opto-electronic position sensors and laser beam should be utilized (MacLeod and Chiarella 1994). All-weather testing was conducted and indicated that the system was effective for work crew protect. However, if it is applied in tunnels under late night tunnel washing condition, human drivers would be at risk due to other drivers' drunk and/or drowsy driving.

Literature review of previous studies shows that there is not much existing research related to autonomous truck in work zones, not to mention examination of human-AV truck platoon interaction in work zones. How human drivers react to autonomous truck platoon when encountering work zone is a question unanswered. In order to investigate this question, Smart Work Zone Deployment Initiative (SWZDI) funded a project on autonomous vehicles in work zones. This project is conducted in a federated transportation simulator system, a safe and high efficient environment, and quantitative measures along with qualitative survey response are collected. Driver behavior, drivers' understanding of truck platoon signage and preference, and effectiveness of education are captured. Statistical techniques are utilized to model and analyze data and survey response.

2.4 PSYCHOPHYSICS AND BIOFEEDBACK UTILIZATION

Biofeedback is a mind-body technique for mental and physical intervention (Frank et al. 2010). Individuals learn to modify physiology to improve physical, mental and spiritual health, like a physical therapy. Biofeedback training provides operant conditioning and feedback learning to guide user behavior to better direction, and psychophysiological psychotherapy is useful for stress component treatment. According to Frank (2010), there are five levels of efficacy: 1, not empirically supported; 2, possibly efficacious; 3, probably efficacious; 4, efficacious, and 5, efficacious and specific.

Biofeedback is divided into two modalities: peripheral and central. Peripheral biofeedback is based on electromyography, electrodermal response, heart rate, temperature, or blood volume pulse. Central biofeedback is based on electroencephalography neurofeedback. Both of them could be used to improve concentration and attention and lower anxiety and disruptive mental chatter (Pop-Jordanova & Demerdzieva 2010).

There are six types of biofeedback products in the current market: 1. eye tracking; 2, Facial Expression Analysis (FEA); 3, Electrodermal Activity (EDA)/Galvanic Skin Response (GSR); 4, Electroencephalography (EEG); 5, Electrocardiogram (ECG/EKG); 6, Electromyography (EMG). Other than the biofeedback products, psychophysiological psycho data could be obtained via survey and Application Programming Interface (API), as shown in Figure 2.4.0.1.

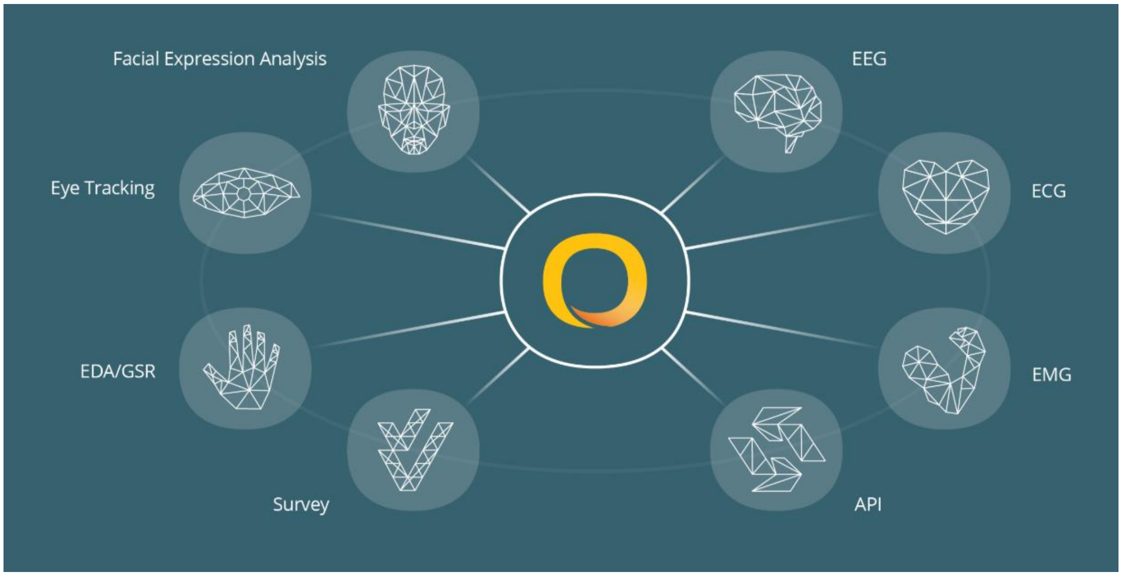


Figure 2.4.0.1 Biofeedback Products (imotions.com 2019)

Eye trackers track eye or pupil position and movement to access visual attention, and know where a user's focus is at any given time. There are three types of eye trackers. Screen-based eye trackers tracks where users look on any screen; eye tracking glasses are wearable, and understand how respondents view and interact in the real, dynamic world; VR tracks attention in Virtual Reality. FEA assesses emotions via facial coding system. EDA/GSR measures stress level via finger tips with heightened skin conductance, as shown in Figure 2.4.0.2. EEG is a head band or a set of head ware which records brain waves, as shown in Figure 2.4.0.3. ECG/EKG measures heart rate activity and EMG measures muscle movements. Devices of ECG and EMG could be shared, as shown in Figure 2.4.0.4.

Figure 2.4.0.2 Example of EDA/GSR Device (imotions.com 2019)

Figure 2.4.0.3 Example of EEG Device (imotions.com 2019)

Figure 2.4.0.4 Example of ECG and EMG Device (imotions.com 2019)

2.4.1 Biofeedback for Stress Levels

Stress level can be determined based on the analysis of Galvanic Skin Response (GSR) and speech signals (Kurniawan et al. 2013). In Kurniawan et al.'s paper, they investigated classification techniques to automatically determine stress based on GSR and speech. When a person's stress increased, his/her respiration rate increased, which appeared to increase pitch or fundamental frequency during voiced section. Physiological signals could be detected by sensors such as GSR. Sensors detected electrical activity on the scalp, blood pressure or blood volume pulse (BVP), or resistance and output voltage of finger tips.

Stress could be assessed and mitigated through peripheral biofeedback. A training for sport performance was conducted for children with mental disorder (Pop-Jordanova & Gucev 2010). There were four groups of children: healthy school children (HSC) (control), children with cystic fibrosis (CF), and general anxiety (GA) and attention-deficit–hyperactivity disorder (ADHD). Within each group, there were 30 children from each gender. Psychological characteristics was evaluated on Eysenck Personality Questionnaire (EPQ). The time used to finish up the game, heart rate, and biofeedback score measured. Results showed that the ability to learn stress mediation was correlated with age. All three groups of children with mental disorder had significantly lower relaxation levels when compared to healthy controls. Relaxation was more difficult for children with GA or ADHD, and easier for children with CF.

Villarejo et al. (2012) designed and built a stress sensor based on GSR. The sensor was controlled by ZigBee (an internet of things solution) wirelessly. The sensor was validated through a set of tests, with 16 adults (eight male and eight female) participating. The tests required certain degree of effort, for example mathematical operations and breathing deeply, and the result turned out that success rate was 76.56%.

Other types of stress assessment and mitigation were conducted. By using electrophysiological biofeedback, NASA measured space motion sickness and built stress training simulators (Gaudeau et al. 2005). Girzadas et al. (2009) to assess stress of residents and medical students learning to play different roles in an airway flight crew, and the results indicated that stress level were not affected by roles.

2.4.2 Biofeedback Utilization in Transportation

A biofeedback product called MoodWings was developed and tested for effectiveness (MacLean et al., 2013). MoodWings is a wearable butterfly that mirrors a user's real-time stress state through actuated wing motion. It functions as an early-stress-warning system, and a physical interface through which users could manipulate their affective state. The research hypothesis was MoodWings would help users to calm down and perform better during stressful tasks. The hypotheses were tested on driving. The results showed that 1) users drove significantly more safely with MoodWings; 2) users experienced higher stress levels (physiologically and self-perceived); 3) users were enthusiastic about MoodWings, expressing several alternative contexts in which they would find it useful. Future study of MoodWings was to design implications for building externalized manifestations of real-time affective state.

In order to reduce vehicle crashes, a study on detecting fatigue of driver was conducted (Bundele & Banerjee 2009). Oximetry pulse measures the oxygen saturation of blood and the changes in blood volume in the skin. A wearable computing system measuring oximetry pulse through skin conductance was utilized to measure mental fatigue and drowsiness. Driving fatigue levels were measured twice, pre and posting driving, producing a photoplethysmograph. Results showed that reducing oxygen supply to the driver while driving causes driver fatigue, and if the oxygen concentrations supplied changes during the drive time, driver would be less fatigue.

Currently, most of psychophysics research focuses on training and treatment, and very few studies are related to transportation, not to mention work zone safety. Among the ones on transportation, they emphasized the perspective of drivers only, without considering the driving environment or any new design. In the autonomous truck platoon in work zones project, biofeedback devices are utilized to measure drivers' stress level, and further interpret the design of signage, impact of truck platoon, and effectiveness of education.

2.5 FACTOR ANALYSIS

Factor analysis is a statistical approach of describing variability among observed variables, finding out correlation among them, and identifying latent factors (Harman 1976). Latent factors are potential unobserved variables, and usually the number of latent factors is less than the number of observed variables. This method helps to divide a large dataset into small ones, making data more understandable and interpretable.

Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) are two commonly used factor analysis methods. EFA uncovers correlations of variables and determines the underlaying structure in a dataset, and CFA deals with the relationship between observation and latent variables, and determines if data fits a hypothesized model from prior theory or assumptions. EFA has advantage of freedom of structuring, which CFA confirms the structure and figures out the estimated coefficients of variables in the structure. EFA acts as a prior step of structural equation modeling (such as CFA) (Ford et al. 1986; Brown 2014).

The application of factor analysis in transportation dates back to 1950s (Garrison and Marble 1964). In 1963, Garrison and Marble (1964) studied the connectivity of transportation network using factor analysis matrix and identified structural relationship within an air transportation system. Factor analysis is also applicable in transportation project management (Nguyen et al. 2015), and risk management (Ackerley et al. 2010).

A behavioral study was conducted in Canada by collecting survey questionnaire response from 2,000 phone interviews of Ottawa-Carleton residents, figuring out eight latent factors from 47 observed variables, concluding that bus service information was the most important factor in public transit ridership (Syed and Khan 2000). A similar study related to public transportation choice of commute was done, using the survey response collected by Regional Transportation Authority in Northeastern Illinois, identifying six factors from 23 questions related to daily travel, and finding out the importance of publicity in attracting new riders, marketing messages focusing on lower stress level and higher time efficiency due to transit use (Popuri et al. 2011).

Factor analysis is commonly used in transportation network analysis, public transit attitude surveys, and transportation project management. However, in literature reviewed, data was analyzed separately, and no previous research linked survey response to driving behavior or performance. This dissertation proposes factor analysis that adopts EFA as prior step and build prior models, then runs CFA to confirm the results from EFA and researcher's assumptions. Data of transportation simulator study, post-simulator survey, field study, and field survey response are combined together for the first time through factor analysis approach.

CHAPTER 3: METHODOLOGY

3.1 SIMULATOR STUDY

A simulator is a machine that provides a realistic imitation of real-world scenarios. In transportation engineering, simulators could be used to represent the operation of vehicles, aircrafts, or other complex systems. Transportation simulators are currently developed for a variety of uses including research purposes, training, and technical analysis, as they are able to provide safe, cost-effective and customized scenarios, reduce risk of field study (Decina et al. 1996), and enable tests to be under control (Kaptein et al. 1996). ZouSim lab in University of Missouri-Columbia built its own federated simulator system, which is a full suite of simulators including car, truck, bike, pedestrian, wheelchair, and e-scooter simulators. The simulators are connected and communicating with each other via a virtual server wirelessly. The truck simulator was set up using a Volvo heavy truck, and the car simulator was a half cab of Corolla sedan, both of them were equipped with Logitech steering wheel, acceleration and brake petals. Scenarios developed via Unity and C# programming, calibration and validation using reality were conducted and are continuing. Psychophysical and biofeedback devices are utilized along with the simulators. All ZouSim simulator studies were followed by a post-simulator survey to obtain demographic information and qualitative measures.

There is a diverse range of disciplines and areas utilizing transportation simulators. Weir and Clark (1995) categorized applications of driving simulators into three types. The first type is related to roadway and environmental variables, including geometric design, roadway texture, color of marking, and other relevant variables (Weir and Clark 1995). The second type of application is related to drivers, including age, gender, education, health status, medical conditions, personality, and other features. The third type is related to vehicles, including vehicle type, height of wheels, display format, field-of-view, and other vehicle conditions. Studies conducted by ZouSim so far are related to roadway designs.

There are two important factors in simulation studies, simulator validity and simulator fidelity. Simulator validity refers to specific research questions and the congruity between simulator configuration and research questions, how simulator tests the objective questions and how they match with each other, but not to the simulator itself in general (Kaptein et al. 1996). Simulator fidelity refers to how close the simulation is to reality, or how realistic the simulator is (van Leeuwen et al. 2015). Fidelity may have impacts on validity. Higher fidelity may result in more normal driving behavior and thus enhance validity. However, if fidelity is too high, it may make scenarios too unique to be adopted into other circumstance, may result in worse simulator discomfort (Dziuda et al. 2014), and even distract participants from the original task (van Leeuwen et al. 2015).

Previous researchers defined driving simulators into different classifications based on physical and functional characteristics (Weir and Clark 1995), set-up equipment (Kaptein et al. 1996), or cost (Blana 1996). As ZouSim simulator equips with large projection screens, vehicle cab, realistic base, vibration simulator under car seat, and steering wheels, and it promotes legitimate fidelity, it fits within mid-level simulator no matter from characteristics based, equipment based, or cost based classifications.

3.1.1 Simulation Sickness

Theories had been developed by researchers over the years, and four of them are widely accepted. Sensory conflict theory claims that the sickness symptoms are caused by the conflict between what the sensory systems feel and actual motion (Oman 1990; Reason and Brand 1975). Postural instability theory argues that since sensory systems have been in an unusual as normal status, the tendency of maintaining postural stability breaks the balance and causes sickness

(Riccio and Stoffregen 1991). Eye movement theory proposed that when errors occur in two specific eye movements, sickness symptoms would arise (Ebenholtz 2001). Evolutionary theory suggests that human bodies need time to adapt to mode shift (Treisman 1977). Previous research also indicates that it is easier for elderlies to get simulator sickness than young adults (Brooks et al. 2010). To reduce simulator discomfort, ZouSim equips with adjustable air conditioners, fans, changeable lights, and provide water and mints to participants.

3.1.2 Sequence Bias and Learning Effects

In any simulator study, there is a possible sequence bias or order effect because the first scenario that a participant encounters can act as an anchor, affecting subsequent encounters (Perreault 1975). As participants go through the trials, they would be more familiar with the simulator, which may shorten their elapse time of later trials, even raising preference on designs which they feel more comfortable with, which is subjective. One way to control for this bias is to randomize the test order.

Even though simulator participants encounter scenarios in a randomized order, which eliminates sequence bias of configurations being tested, when the measure is repeated, the overall performance from the second trail may be different from the first trial. Statistical analysis of learning effect is presented in Chapter 5, Section 5.3.

3.1.3 Human Factor

Participant Recruitment

Unlike other aspects of civil engineering, transportation simulator research is mainly human subject tests. In order to reduce bias due to demographic differences, recruitment of participants is open to the public. A balance between genders, age groups, occupation, and residency is sought. In transportation simulator studies, it is commonly accepted that 30 participants are recruited for data collection, with consideration of time, budget, and statistical reasons.

According to Hogg and Tanis (1977), a sample size of 30 could be regarded as the boundary of large sample, and the sample could be assumed as normal distribution.

However, as simulators are parked in ZouTrans Lab, and requires participants coming to campus, it is very possible that the demographic distribution of participants is not balanced. Most of the recruitment occurred through network of students, faculty and staff, and participants are mostly friends and family members. Younger age group (18 – 40) takes up a large proportion. Most of participants are urban residents, as the simulator tests take place in City of Columbia.

One possible way to improve participant recruitment process is to increase accessibility of driving simulators. Simulators could be placed on movable bases and taken out to public areas for short term data collection. As the public are able to see the simulators, their interest may raise due to easy accessibility. With the amplified exposure, the chance of obtaining higher sample size and balanced demographical groups would increase.

Hosting Manual and Host Training

In simulator study, participants may react to different hosting languages, which would result in different driving behavior and data collection. To eliminate this bias, standardized hosting manual is required for each trial. All of the trial hosts should be trained to follow the wording and body language exactly as it is instructed in the standardized hosting manual.

Legal Issues

Simulators are emerging to study safety and human factors in transportation as well. Due to recent implementations, the knowledge of the legal issues pertaining to these topics is limited. There are potential legal issues pertaining to the interaction between participants and researchers, including relevant privacy issues and proper notice to the participants in form of waiver documentation.

During the simulator trials, participants contribute personal information and their response to the test scenario. Similarly, the researchers need to provide relevant information to the participants. In this dissertation, the team focuses on the exchange of information between participants and researchers involving privacy issues and associated documentation.

Privacy

To satisfy the sample size regarding age, gender and other factors, the tests using driving simulator not only need to collect the reaction or response to alternative design or specific scenarios, but also need to collect related personal information, even videotape the whole process of testing. Therefore, privacy policy and information access control are essential to transportation simulator tests.

Regarding privacy policy, it is not just a dispensable reminder for researcher, but a heavily valued legal issue in every business. There are laws in place to protect the privacy of individuals that researchers must always maintain. Privacy management in transportation simulators mainly contains four parts: content of personal information collected, purpose or use of the personal information, security of the information, and accessibility of personal information.

The first part of privacy policy issues involves informing the participant about the personal information that will be collected during the transportation simulator test. In the case of Intervention v. Blizzard Entertainment, the suit claims that while a game player is online, the software transmits the gamer's name and e-mail address without permission or knowledge (Knefel 1998). In other words, all the information collection and usage related to research should be informed to participants. The second part involves the purpose or use of the collected information. The conducting research may not be used for other purpose, otherwise it is forbidden. In the case of Lahr v. National Transportation Safety Board (NTSB) (Lahr v. National Transp. Safety Bd. 2006), to assist in the crash investigation of Trans World Airlines Flight 800 in 1996, Boeing voluntarily provided information to the Central Intelligence Agency and NTSB. Boeing flight license and privacy policy state that data could be used in proprietary flight simulators for flight training, engineering and other commercial purposes, so this action did not breach any law.

The third aspect of privacy management is the security of collected information. Hackers target profitable companies as well as valuable research data; therefore, researchers or operators of transportation simulators should increase the awareness of privacy protection. The fourth part incorporates who could access personal information. Access needs to be limited to participants and internal researchers. In the case of Lahr v. NTSB (Lahr v. National Transp. Safety Bd. 2006), the plaintiffs requested to publicize the data and information gathered by Boeing's flight simulators. Boeing rejected and claimed that the data was confidential and proprietary, and the court ruled in favor of Boeing. Law enforcement and government agencies are parties with accessibility to one's person information as well. One such example involved the missing Malaysian airline flight. Malaysian investigators recovered data from a flight simulator found in the home of the missing airplane pilot. Authorities sought to find more clues about the final destination of the plane (Harlan et al. 2014).

In applying this legal knowledge to the ZouSim simulators, privacy management should be of high importance. Participants need to be informed of the data collection, the use of information, how it would be protected, and who might have access to their personal information.

Notice (Waiver Documentation and Informed Rights)

Communication between participants and researchers includes the information collected from the participant and relates to the notice offered to the participant. Proper notice to the participants includes providing waiver documentation and informing them of their rights.

In order to protect both participants and researchers, a well-constructed waiver document could be beneficial to both parties. It could serve as a way to acknowledge the rights and privileges of the participants and make aware a variety of potential risks that participants may be subject to, which could include privacy issues, stressful environments, or other possible risks. A waiver document is designed to allow researchers to be honest of foreseeable risks and share a portion of liability among the participants. However, the document does not give researchers absolute immunity, as there are laws in place which govern certain limitations of waiver forms. Missouri (MO) has both statutes and common laws in place that determine certain instances in which a waiver form could be deemed invalid. These include improper location of critical phrases, wrongful intent of the waiver, and lack of clarity.

The case of Alack v. Vic Tanny International of MO Inc. involved an injury that took place at a health club facility. The plaintiff signed a waiver form in which he agreed to assume liability for a fault of his own behavior. However, the document was worded in a way that he would also assume liability if the injury was a result of negligence by the facility. The court deemed the waiver document invalid due to a lack of clarity surrounding a user’s liability and the negligence enacted by the health club facility itself (Alack v. Vic Tanny International of Missouri Inc. 1996). This case also stated that such a contract should not be upheld due to the location of such crucial wording.

Understanding and informing the participants of their rights is another important aspect of simulator research. The Institutional Review Board (IRB) is a nationally recognized entity which aims to protect human participants involved in research projects across university campuses. The IRB has rules in place for informing of the participants of their rights, and it even requires certain elements to be included in a waiver document (IRB 2015). According to IRB, there are eight required elements in the waiver document: statement of research, description of risks, description of benefits, alternatives, confidentiality, compensation, contact information, and statement of voluntary. These elements should be addressed in the simulator waiver.

Careful consideration should be taken in developing a waiver document for the simulator research. The waiver can be used by the research group to share some of the burden of liability with the participant, so long as the document is clearly written, honest in nature, and easily understood by the participant. The waiver should also be fair to the participant by informing them of their rights and contain the eight elements required by the IRB.

3.2 EXAMPLES OF SIMULATOR STUDIES APPLICATIONS

Other than work zone research, transportation simulators can be used for geometric design such as J-turn and diverging diamond interchange, signage design, even non-motorized options, for example wayfinding system development in airport for wheelchair users, e-scooter and bicycle wayfinding and pavement marking. Federated simulator systems can provide more options by connecting simulators together and interact with each other, for instance pedestrian interaction with autonomous vehicles and human driver interaction with autonomous truck platoon.

3.2.1 Evaluation of Bicycle Wayfinding and Detection Marking in Simulator

Urban and arterial work zones focus mostly on motorized vehicles. Although FHWA work zone safety rule requires that non-motorized travelers are accommodated in work zones (FHWA 2012), it is very often that they do not properly guide pedestrians, bicycles, wheelchairs, and other types of non-motorized transportation modes. Sometimes sidewalks or bike lanes terminate at the work zone, forcing travelers into the traffic lanes, which lower the efficiency of traffic flow and increase the risk of crash. Investigating proper wayfinding markings and signage helps to improve work zone navigation.

The demand of effective bicycle wayfinding and detection marking systems is rapidly emerging, as safety is first priority of bicyclists. Finding the best way to set signage and mark the routes is urgently required. In order to test the visibility and sufficiency of new designs, a bicycle simulator study was conducted, providing a safe and cost-effective environment. This study could be used as an example to be extended for work zones accommodation and provide guidance for non-motorized transportation modes.

Two new sets of wayfinding signage were proposed, and compared with MUTCD standard signage, as shown in Figure 3.2.1.1. Instead of putting signage on curb or sidewalk, the new signage was marked on pavement to improve visibility and catch bicyclists' attention easier. Three types of detection marking designs were proposed, and compared with MUTCD marking, and MUTCD marking plus signage, as shown in Figure 3.2.1.2.

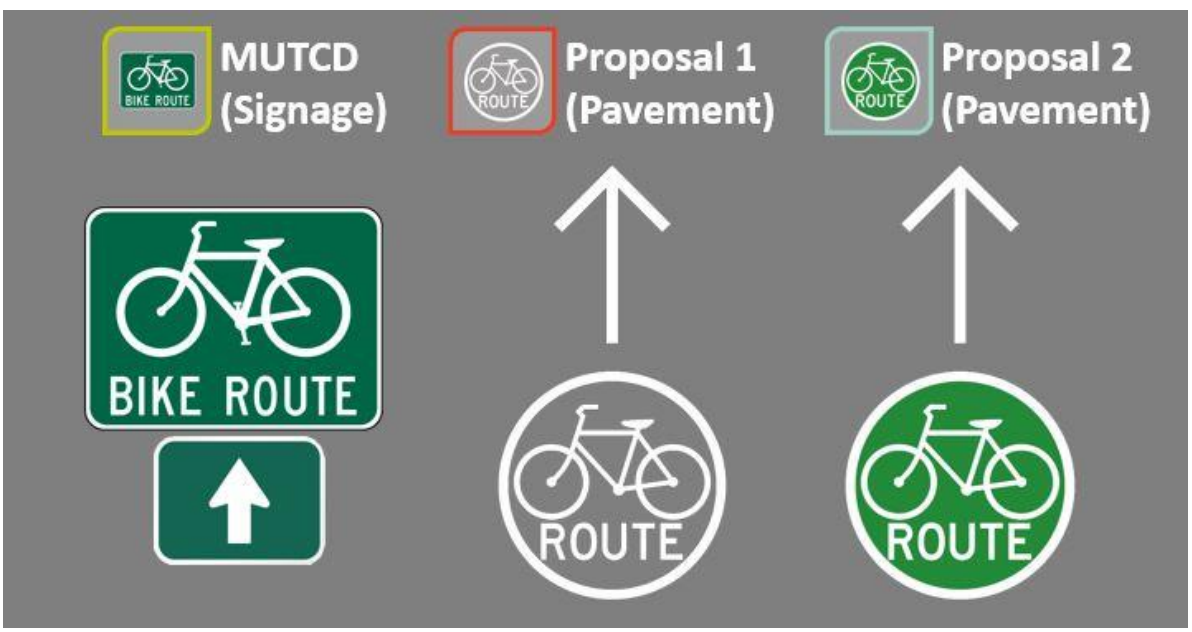


Figure 3.2.1.1 Wayfinding Signage and Pavement Marking

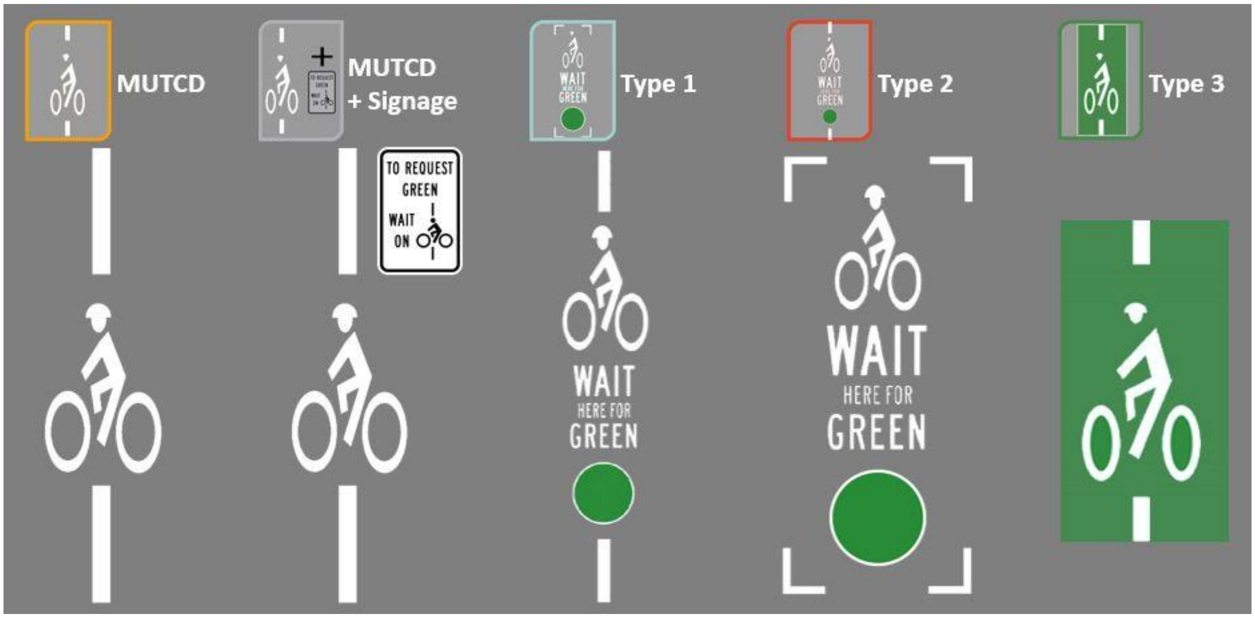


Figure 3.2.1.2 Detection Marking

The routes and transitions were simulated in Unity, and speed and motions of bike were transformed to digital signals by a set of physical equipment attached to the bike. Participants were asked to ride three runs and their behavior were recorded in videos. An example of a participant taking simulator trial is shown in Figure 3.2.1.3.

Figure 3.2.1.3 Example of a participant taking simulator trial

Each wayfinding designs were tested in one run, and each type of detection markings were tested twice in total. The trials were run in order, and the experiment design is shown in Table 3.2.1.1. The first run tested MUTCD wayfinding signage, with ten wayfinding intersections and three detection intersections. The second run tested Proposal 1 wayfinding pavement marking, with ten wayfinding intersections and three detection intersections. The last run tested Proposal 2 wayfinding pavement marking, with ten wayfinding intersections and four detection intersections. The total duration for the simulation experiment was around twenty-five minutes. After the runs, participants were asked to fill in an online post-simulator survey. The survey collected participants' understanding of designs and preference, and checked for their experience and feelings of simulator sickness.

Table 3.2.1.1 Experiment Design

	Wayfinding Type	Number of Intersections	
		Wayfinding	Detection
Trial 1	MUTCD	10	3
Trial 2	Proposal 1	10	3
Trial 3	Proposal 2	10	4

The initial number of participants recruited was 30, with 25 participants completing both the simulator runs and surveys. Among the other five participants, two felt sick during the experiment and were not able to complete the runs, but completed the survey afterwards, and three did not participant due to schedule conflict. Objective data was extracted from video recording of the trials, and subjective data came from survey response.

Overall results are shown in Table 3.2.1.2. Time stamps from videos showed that the average running time for MUTCD wayfinding signage was the longest, and the average running time for Proposal 2 was the shortest. Total number of incorrect intersections turning for MUTCD was the most, and for Proposal 2 was the least. For hand signals before seeing wayfinding signage or marking, participants made hand signals the latest and closest to MUTCD signage, and the earliest and furthest to Proposal 2 marking, as shown in Table 3.2.1.3. Average waiting time on detection marking of MUTCD pavement marking plus signage was the longest, and waiting time of Type 1 and Type 2 were the shortest, as shown in Table 3.2.1.4.

Table 3.2.1.2 Overall Results

	Average running time	Total incorrect transitions
Trial 1	08:15.9	48
Trial 2	08:12.3	24
Trial 3	08:08.6	15

Table 3.2.1.3 Results for Signage

Signage	Reaction time (s)	Distance to signage (ft.)
MUTCD	0.53	17.54
Proposal 1	2.14	41.39
Proposal 2	2.94	47.96

Table 3.2.1.4 Results for Detection Marking

Detection Marking	Incorrect detection (out of 48)	Average waiting time (s)
MUTCD	28	15.08
MUTCD + Signage	29	15.48
Type 1	23	13.10
Type 2	23	13.10
Type 3	27	14.69

The overall objective data result showed that Proposal 2 wayfinding pavement marking performed the best, because of the shortest elapsed time, least incorrect intersection turnings, and earliest hand signal. Both Type 1 and Type 2 detection marking designs have the best results due to the shortest waiting time.

The survey asked participants about the marking visibility, effectiveness, clarity, and preference. Results showed that Type 1 marking scored the highest in all these four aspects, and MUTCD markings was the worst.

Both objective and subjective results showed that wayfinding system in Proposal 2 and Type 1 detection marking turned out to perform the best and are most preferred. However, sequence bias may influence the reliability of result, because participants preferred the Proposal 2, which was tested in the last run, and participants may have learned and got used to scenarios. Although experiments indicated that Proposal 2 wayfinding design and Type 1 detection marking were optimal, engineering concerns would be considered in practical designs. This study could be further expended and adapted into work zone wayfinding and signage marking to guide non-motorized transportation mode users.

3.3 SYNERGY OF SIMULATOR AND FIELD STUDIES

Simulator study is now playing a more significant role in transportation, as it provides safe testing environment and high efficiency of testing different configurations. More importantly, simulator offers the opportunity to test for new designs do not exist in reality, or extreme driving conditions that are risky even fatal in real life. However, results from simulator studies could only examine for relative validity of multiple designs and configurations, without being able to test for absolute validity.

Field study is more transitional and commonly used method in transportation, as it collects data in reality, and the driver behavior captured is closest to real life observation. However, field study has limitations, such as lack of equipment, risky testing environment, and constrained by weather condition. Besides, under most of circumstances, written qualitative response is not possible to obtain as drivers are not accessible in continuously flowing traffic.

Considering both advantages and disadvantages of simulator and field studies, it is obvious that the synergism of them is of high importance. Conducting field study before simulator trials helps with validation and calibration for simulator configuration and scenario design. Accompanying field study after simulator trials provides reality support for simulator results. Matching measurements of effectiveness (MOEs) is one way to collaborate the studies, so that for important measures, both relative and absolute validity are presented for comparison. However, in some cases, match of MOE is not realistic due to limitation of technology. For example, in a simulator study, capturing the distance between participant driven vehicle and a mobile work zone when the vehicle merges to passing lane is doable, but in a field study, a vehicle may have merged very early, and the distance of out of detection range or field of view of a recording camera.

Although simulator studies provide only relative validity, it is feasible to combine it with absolute results from field studies, to examine the relationship and determine if they could explain each other well. Statistical analysis of simulator and field studies synergy is presented in Chapter 5, Section 5.1 and Section 5.2.

3.4 PSYCHOPHYSICS UTILIZATION IN TRANSPORTATION SIMULATORS

3.4.1 Eye Tracking

Eye tracking devices are more widely deployed in recent studies. In simulator studies, an eye tracker tracks the movement of a participant's pupil, capturing the frequency and time of participants looking at specific spots. This is useful to examine designs involving position configurations, signage, or multi-media information platform for example apps on hand-held devices. Also, eye tracker can act as an indicator, indicating that if a participant captures an object in a disability glare test.

3.4.2 Stress Level Measurement

Although field study captures driver behavior in reality and presents absolute validity, drivers' mental status and pressure level were not able to be obtained during field study. Simulator study makes it possible by incorporating psychophysical devices in trials.

Among the aforementioned six types of biofeedback products, GSR, EEG, and ECG are able to measure stress level. Based on the flexibility and intrusiveness, GSR products may be useful in transportation simulator trials, as it is attached to fingertip and is easy to install and remove. Due to the relative smaller size, GSR may have less impact on driver's natural movement and driving behavior.

3.5 IMPORTANCE OF EDUCATION

The disruptive technology makes a great contribution in transportation, increasing efficiency, reducing risk, and improving safety. Meanwhile, as technology updates quickly, public education becomes more and more important, because public may misunderstand new signage, new geometric design, or new configuration, which would cause noncompliance, congestion, or even crash.

In a recent autonomous truck simulator study, Schoelz (2019) found that the driver behavior before and after explaining a signage displayed at the back of a truck platoon were different. Before education, participants did not understand the meaning of signage, and did not know how to react to a forming truck platoon, so they stayed in between two trucks. After explain to them that the signage meant trucks were trying to get into a group and go together, driver behavior of participants became more uniform and merged to passing lane, allowing trucks to platoon.

Novelty effect played an important role in public's reaction to new devices and configuration as well. From the noncompliance rate of AFAD field study, it could be seen that some drivers misunderstood the STOP sign on CMS. Instead of stopping and waiting for a SLOW sign, they regarded it as a regular stop sign, stayed for a couple seconds and restarted, which was risky as traffic was coming from the opposite direction sharing one lane. In the study of green lights on TMAs, survey comments mentioned that the green light was confusing as green means go intuitively.

Enhancing public education is essential. When new designs are deployed, they need to be understood by the public, so as to improve safety and efficiency and reduce crash and congestion.

3.6 SURVEY DESIGN

There are many ways of surveying to obtain descriptive data. One common method is to adopt preestablished questionnaire if the information to be obtained is genetic and commonly asked. For example, employee satisfaction survey, censes, or customer satisfaction survey. In transportation engineering, one example of preestablished survey is the customer satisfaction survey in MoDOT quarterly tracker (MoDOT 2019). Questions in the survey are predesigned to be a template, and every time the survey is handed out, the contents are the same as the last round, until the template is modified. Another commonly adopted questionnaire is the simulator sickness questionnaire (SSQ) by Kennedy et al. (1993), as shown in Appendix A. The SSQ contains 16 questions, asking for simulator participants' symptoms during and after simulator trials, which is widely used in diagnosing the simulator sickness severity of participants. The SSQ was attached in all post-simulator surveys in ZouSim simulator studies.

Since the surveys administrated in projects of this dissertation were field surveys and post-simulator surveys, they were specifically designed to ask questions related to the field study and simulator trials, and needed to be initiated by researchers, instead of adopting a general questionnaire. Researchers need to be careful about the language used in survey design, as poor worded questions would affect the validity of survey response. In his book, Frey (1989) listed some wording methods that should be avoided, including loaded questions, use of inflammatory words, slangs, and etc.

3.6.1 Understanding of New Designs

When designing surveys related to new configurations of traffic control devices, or new types of signage, one essential question is to ask if survey respondents understand the meaning of the new design. One simple way is to attach a picture of the design and provide multiple choice options, and ask the survey respondents to select the meaning. For example, in Figure 3.6.1.1, an AFAD with a stop paddle and a CMS showing "STOP" is provided in the survey, asking survey respondent to circle the meaning. Options in multiple choices should be related to the design, and should also consider putting some possible wrong understanding.

Figure 1b

2. What is the meaning of Figure 1b?

a. Wait if STOP sign indicated.
b. A regular STOP sign, make a full stop and go.
c. This is a traffic signal, stop on the red light.
d. The device makes no sense.

Figure 3.6.1.1 Example of Question Related to Understanding

3.6.2 Rating Scales

When surveys are trying to capture respondents' opinions on specific features of configurations, rating is one simple way to achieve. There are two commonly used scales in rating, one is the Likert scale (Likert 1932), which is commonly used for agree and disagree statements, and the other one is from one to 10.

When asking survey respondents to rate, it is important to clearly indicate the corresponding meaning of the numbers. When using Likert scale numerically, usually smaller number indicates less agreement, for example in five-point Likert scale, one means strongly disagree and five means strongly agree. Similarly, in one to 10 scale, one means lowest score and 10 means highest score.

For example, in Figure 3.6.2.1, simulator participants were asked to rate the clarity, visibility, safety and efficiency features of the four traffic control configurations, with snapshots taken at the same distance from AFADs/human flagger, and the same angle of view.

6. **Please rate all designs from a scale of 1 (lowest) to 10 (highest) with respect to the following attributes:**

	Figure 1a 	**Figure 1b** 	**Figure 1c** 	**Figure 1d** 
Clarity				
Visibility				
Safety				
Efficiency				
Comments				

Figure 3.6.2.1 Example of rating features

Figure 3.6.2.2 is another example of Likert scale. It is straightforward to ask for agreement and disagreement when the respondents are able to circle or pick options. Converting agreement to five-point scale is often used in phone interview, when survey respondents have to tap keypads on their phone. Five-point conversion is an essential step in Likert scale data analysis.

8. I felt like I was actually there on the highway.

[ ] Strongly agree [ ] Agree [ ] Neutral [ ] Disagree [ ] Strongly disagree

Figure 3.6.2.2 Question for Simulator Fidelity

3.6.3 Demographic Information

Most of surveys and questionnaires collect demographic information of respondents. The most commonly asked demographic questions are age or age group and gender. Household income information is collected in surveys related to socioeconomics, as well as transportation questions related to the public, such as transit fare, gas tax, and public ridesharing programs.

In surveys of projects discussed in this dissertation, residency and regular vehicle types are asked along with age and gender. The reason for asking residency was because the configurations of urban and rural roadways are very different from each other, which may lead to different driving habit and cause different driving behavior in a less familiar driving environment setting. As simulator participants drove the car simulator, which was a compact size sedan cab equipped with steering wheel, acceleration and deceleration paddles, participants may not be comfortable with the vehicle size and height, and the driving behavior may be different.

3.6.4 Attaching Figures in Surveys

When figures are attached in survey questions, they should be clear and no fuzziness. If the figures are for a field survey, using real photos are preferable. If the figures are for a post-simulator survey, then a high-resolution snapshot with reasonable angle and scale would be fine. Try to demonstrate exactly what the survey respondents would encounter and see when they drive through. Sometimes due to novelty effect or lack understanding of terminology, respondents may get confused with what the question is referring to, then circling out or highlighting the exact object or feature is needed. For example, in Figure 3.6.4.1, the question used the word “message board” instead of “Changeable Message Sign” or “CMS” to avoid terminology, and circled it out for clear referring.

Figure 2

7. **To what extent do you believe the message board in Figure 2, as circled in green, is necessary?**

[ ] Strongly agree [ ] Agree [ ] Neutral [ ] Disagree [ ] Strongly disagree

Figure 3.6.4.1 Example of Circling out Object or Feature Referred by Question

CHAPTER 4: SMART WORK ZONE TECHNOLOGIES

The case studies in this chapter illustrate the utilization of transportation simulators for work zone safety improvement, and show how simulator study, post-simulator survey, and field study support and validate each other. Case study one, evaluation of Automated Flagger Assistance Devices (AFADs), was based on stationary work zone at rural two-way two-lane highway. Case study two, evaluation of green lights on Truck-Mounted Attenuators, was conducted on two-way four-lane freeway with low traffic volume, aiming to improve mobile work zone safety.

4.1 AUTOMATED FLAGGER ASSISTANCE DEVICES (AFADS)

A new type of AFAD was developed by the Missouri Department of Transportation (MoDOT) (Figure 4.1.0.1). This AFAD uses STOP/SLOW paddles and flashing red/yellow lights. In addition, a changeable message sign (CMS) was installed displaying a series of four messages, and the whole set of AFAD was mounted on a TMA, which was the innovative design by MoDOT. As shown in Figure 4.1.0.1, the CMS alternates between an image of a STOP sign and the word “STOP” every two seconds during the stopped interval. The CMS alternates between an image of SLOW and the words “Go on Slow” every two seconds during the go interval. The AFAD was built onto a truck-mounted attenuator (TMA) unit in order to provide protection for the AFAD operator in the truck. The truck integration obviates the need to tow and deploy trailer-mounted AFADs. The MoDOT AFAD was tested in the manual operation mode; the flagger in the truck controlled the signals by observing traffic at the end of work zone and communicating with another flagger at the other end of the work zone by radio.

Figure 4.1.0.1 AFAD Mounted on TMA

Flaggers have been playing an important role in traffic control for a long time, as they guide and direct vehicles on the highway, and often, through work zones. Flaggers are trained professionally to display uniform gestures for traffic guidance using signaling devices. Richards and Bowman (1981) examined the effectiveness of flagger gestures and signals and found that some signals are more effective than others. They also validated the importance of using flaggers. Flaggers are exposed to safety risks, as they may be hit by oncoming traffic when drivers are not aware of the presence of workers or are not able to come to a full stop when approaching the work zone.

This chapter documents the results from a study to evaluate AFADs using four different types of techniques: field monitoring, field survey, driving simulator, and post-simulator survey. Detailed driver behavior measures were used for the first time in this study to compare the effectiveness of human flaggers versus AFADs in the United States.

4.1.1 Field Study

Two major tasks of the project are to conduct field and simulator studies to verify AFAD effectiveness and to study driver behavior. The project includes three phases involving the use of an AFAD: a field test with Changeable Message Sign (CMS), a simulator study (both with and without CMS), and a tentative field test without a CMS. This chapter describes the field test with CMS. The third phase was deemed unnecessary as the use of a CMS was found to be desirable by the first two phases.

Field Set up Plan

Phase one focused on comparing a MoDOT STOP/SLOW AFAD mounted on a TMA against a human flagging system using field data. Video cameras, speed radar guns and delineators were deployed to collect data measurements. Driver performance and driving behavior at both AFAD and human flagger sides were recorded. These driver performance measures included vehicle approach speed, full stop location, reaction time and other unusual driving behaviors.

The field study plan is shown in Figure 4.1.1.1. The camera was placed on the right side of road, to avoid influencing opposite traffic. To measure the vehicle approach speed, the speed radar was set in front of the video camera without blocking the view of vehicles, delineators, and the AFAD or the flagger. The delineators were placed every 50 feet along the road. There were a total of seven delineators from the stop control on each side of the road. In addition to the driver reaction measures, the camera also recorded traffic information on the road, such as traffic volume, waiting time deviated from video recording, and queue length.

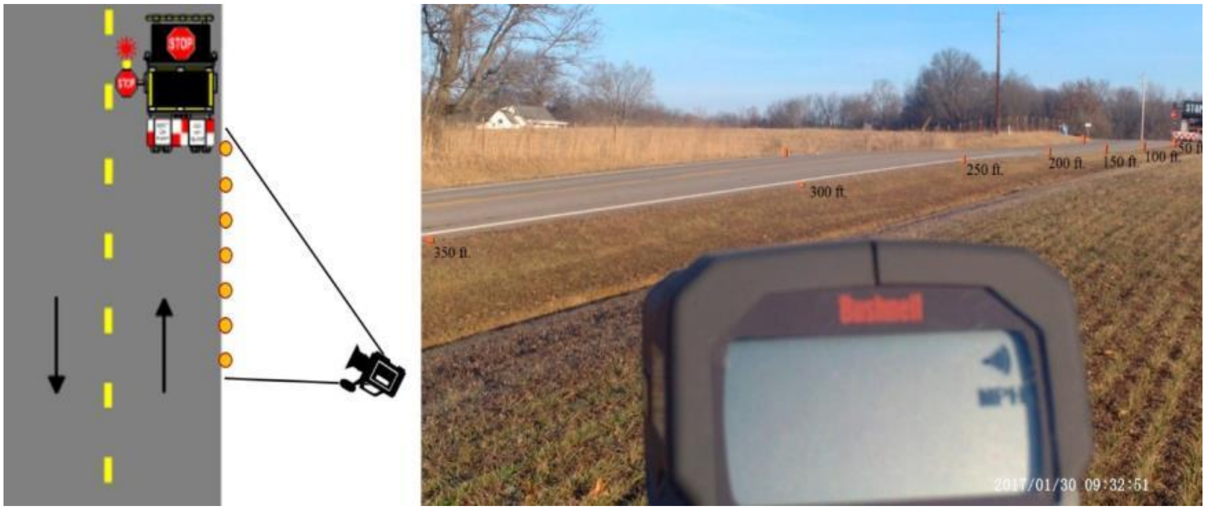


Figure 4.1.1.1 Field Study Plan of Cameras, Radar Speed Gun, and Delineators

Two field data sessions were conducted to collect field data. The first one was on December 20th, 2016, on MO150 in Lone Jack, Missouri. The second one was conducted on January 30th and 31st, 2017, on MO-23 Highway in Knob Noster, Missouri.

First Field Data Collection

The first field data collection was on December 20th, 2016, on MO-150 in Lone Jack, Missouri. MO-150 was a two-lane highway, and the work zone was 2,200 feet long from the AFAD on one end to the flagger at the other end. The annual average daily traffic (AADT) on the road segment was 1,028 vehicles per day, according to the MoDOT Transportation Management Systems (TMS). The work zone layout and descriptions are shown in Figure 4.1.1.2 and Table 4.1.1.1.

Figure 4.1.1.2 Map of MO-150 Work Zone (Google Maps 2017)

Table 4.1.1.1 MO-150 Field Data Collection Information

Location:	MO-150 in Lone Jack, MO
	Two-lane highway
	Speed Limit 45 mph/55 mph
AADT:	1,028 vpd (directional 514)
Length:	2,200 ft. (from the flagger to AFAD)
Duration:	12/20/2016 10:30 AM - 11:45 AM

In the field, one camera, one radar speed gun, and a set of delineators were placed at each work zone end. The field settings and the field views of the cameras are shown in Figure 4.1.1.3. The west end camera recorded traffic and driver reaction to the flagger, and the east end camera recorded activities at the AFAD.

Figure 4.1.1.3 Field Settings on MO-150 Highway Work Zone (Google Maps 2017)

The data collection was conducted from 10:30 AM to 11:45 AM. As the testing was underway, it was discovered that the AFAD was not functioning properly. Therefore, the data collection was aborted. Subsequently, MoDOT changed the AFAD controller from wireless to wired to enhance reliability, and a second field study was scheduled.

Second Field Data Collection

The second field data collection was conducted on January 30th and 31st, 2017, on MO-23 Highway in Knob Noster, Missouri. The work zone was 2,400 ft. long and the AADT value on the road was 2,610 vehicles per day. The work zone layout and information of work zone are shown in Figure 4.1.1.4 and Table 4.1.1.2.

Figure 4.1.1.4 Map of MO-23 Work Zone (Google Maps 2017)

Table 4.1.1.2 MO-23 Field Data Collection Information

Location:	MO-23 Hwy, Knob Noster, MO
	Two-lane highway
	Speed limit 55 mph
AADT:	2,610 vpd (directional 1,305)
Length:	2,400 ft. (from the flagger to AFAD)
Duration:	01/30/2017 09:17 AM – 04:47 PM
	01/31/2017 09:57 AM – 04:29 PM

Data Collection

The data collection deployment on MO-23 Highway is shown in Figure 4.1.1.5. The difference between the deployment and the field study plan was that the north side camera was placed on the left side of road due to topographic constraints. This change had minimal impact because the small volume of opposing traffic did not occlude the camera. On one end of the work zone, there was an AFAD mounted on a truck-mounted attenuator (TMA) truck (Figure 4.1.1.1) with an operator sitting inside the TMA vehicle to control the AFAD remotely. On the other end, there was a human flagger standing next to the traffic lane to show STOP/SLOW paddles (Figure 4.1.1.6). Speed guns and cameras were set up at each side to record approaching speeds of vehicles. On the first day, the AFAD was located on the south side of the work zone, and the flagger was on the north side. On the second day, the locations of the AFAD and flagger were reversed. Thus, each type of flagging was deployed at both ends of the work zone.

(a) MO-23 Work Zone (Google Maps 2017)

(b) South end, first day (AFAD) **(c) North end, first day (Flagger)**

(d) South end, second day (Flagger) **(e) North end, second day (AFAD)**

Figure 4.1.1.5 Field Settings on MO-23 Work Zone

Figure 4.1.1.6 Flagger

Data Processing

Field videos were reviewed and performance data were obtained. Only vehicles that encountered the STOP message were processed; those vehicles that encountered the SLOW display and drove through directly were not processed. The reason for focusing on stopped vehicles was to assess the safety impacts of flagging systems. The number of samples is shown in Table 4.1.1.3. The sample size collected was 334 in total, of which 186 was for AFAD and 148 for flagger.

Table 4.1.1.3 Summary of Field Data Collected

Field Data	Traffic Control Types		Total
Location	AFAD	Flagger	
South End	102 (First Day)	82 (Second Day)	184
North End	84 (Second Day)	66 (First Day)	150
Total	186	148	334

After the field data was collected, the research team reviewed the videos, and conducted the data reduction process. Seven Measures of Effectiveness (MOEs) were defined for data reduction as described below.

- MOE 1: speed of the leading vehicle at 250 ft. from the AFAD/Flagger (Figure 4.1.1.7). The speed was read from the speed gun. However, the speed gun did not display any speeds lower than 10 mph, so researchers estimated speeds less than 10 mph using the speed from the last reading.

Figure 4.1.1.7 MOE 1 Example: Speed of the Leading Vehicle at 250 ft. from AFAD/Flagger

- MOE 2: full stop location (Figure 4.1.1.8). Location of vehicle's front end when the vehicle came to a full stop. The location was the distance from the AFAD or flagger. The distance was determined from the video based on the delineator cones that were placed.

Figure 4.1.1.8 MOE 2 Example: Full Stop Location

- MOE 3: waiting time (Figure 4.1.1.9). Waiting time was measured as the time gap between the time when the vehicle came to a full stop and when the vehicle started to move again after receiving the SLOW indication from the flagger or AFAD.

Figure 4.1.1.9 MOE 3 Example: Waiting Time

- MOE 4: reaction time (flagger/AFAD CMS) (Figure4.1.1.10). Reaction time was measured as the time between when STOP changes to SLOW (paddle for flagger and CMS for AFAD) and when the vehicle restarts. At the time of the field experiment, the SLOW paddle on AFAD and the messages on CMS were not fully synchronized. When the message on the CMS changed from STOP to SLOW, the paddle started to turn, and it took four seconds to finish turning. Drivers appeared to react based on the message shown on CMS. The time lag between the paddle and CMS on the AFAD was corrected after the field testing. For the AFAD, reaction time was measured based on the CMS and not the paddle since drivers appeared to react to the CMS.

Figure 4.1.1.10 MOE 4: Reaction Time

- MOE 5: Noncompliance rate (Figure 4.1.1.11). Noncompliance refers to when a vehicle ignored the STOP sign and was stopped by the AFAD or flagger. If a vehicle came too close to the AFAD or tried to go through, then the AFAD truck would sound its horn. If a vehicle came too close to the flagger, then the flagger stopped the vehicle via gestures. In either case, it was regarded as one noncompliance. Noncompliance rate equals the ratio of noncompliance over the sample size.

Figure 4.1.1.11 MOE 5: Noncompliance Rate

- MOE 6: speed of the 1st following vehicle at 250 ft. from AFAD/Flagger. Similar to MOE 1, the speed was again captured at 250 ft, as shown in Figure 4.1.1.12.

Figure 4.1.1.12 MOE 6: Speed of the 1st Following Vehicle at 250 ft. from AFAD/Flagger

- MOE 7: queue length (Figure 4.1.1.13). The number of vehicles in a queue.

The seven MOEs were extracted and data were grouped by direction (southbound/northbound), flagging type (AFAD/Flagger), and vehicle type (sedan, pickup, commercial vehicle). Since the height of vehicles impacts sight distance, taller passenger vehicles such as SUVs, pickups, and minivans were differentiated and labeled as the pickup category.

Figure 4.1.1.13 MOE 7: Queue Length

<u>Field Data Results</u>

All of the MOEs were recorded and extracted from videos. Differences between MOEs were calculated to compare AFAD and flagger performance, and t-test was run to determine if the difference was significant. Confidence level was indicated by the t-test result, and the effect size was indicated by Cohen's d.

The Student's t-test is a statistical test that compares the means of two groups and assesses if the means are statistically significantly different from each other (Milton and Arnold 1995). T-score is a ratio between difference of two means and difference within each group. A higher t-score means larger difference between two groups. P-value is defined as the probability of obtaining extreme results as observation under null hypothesis. It ranges from 0% to 100%, with lower value more desired to reject null hypothesis, and p-value less than 0.05 is commonly accepted as threshold of statistical significance. Each t-score with corresponding degree of freedom is associated with a p-value. Currently, software such as Microsoft Excel, IBM SPSS, SAS, would present p-values as t-test results instead of t-score. In transportation engineering, two sample t-test is commonly used to compare the effectiveness of two configurations or designs. It is applicable for field data and simulator quantitative data analysis.

Two sample t-test is a good fit for comparison between two groups. The null hypothesis is that the means of these two groups are not different from each other. The calculation of t-score is

$$t = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

where $\overline{X_1}$ is the mean of group 1, $\overline{X_2}$ is the mean of group 2; s_1^2 is the variance of group 1, s_2^2 is the variance of group 2; n_1 is the sample size of group 1, and n_2 is the sample size of group 2. When the variance of two groups are equal, then the t-test is called Student's t-tests. Without this assumption, the t-test is called Welch's t-test (Welch 1947).

Cohen's d is an effect size, which indicates the standardized difference between two means. It is commonly used to accompany reporting of t-test and ANOVA results, as well as meta-analysis. Cohen’s d could be calculated as

$$Cohen's\ d = \frac{|\overline{X_1} - \overline{X_2}|}{\sqrt{\frac{s_1^2 + s_2^2}{2}}}$$

where $\overline{X_1}$ is the mean of group 1, $\overline{X_2}$ is the mean of group 2; s_1^2 is the variance of group 1, s_2^2 is the variance of group 2.

Effect size is defined to be small when Cohen’s d is between 0.2 and 0.5, medium when Cohen’s d between 0.5 and 0.8, and large when Cohen’s d is larger than 0.8 (Cohen 1977). Effect size is only applicable when Cohen’s d is larger than 0.2, because when Cohen’s d is smaller than 0.2, the difference was not statistically significant at 95% or higher.

MOE 1 measured the speed of the leading vehicle at 250 ft. from the AFAD/Flagger. As shown in Table 4.1.1.4, the average approaching speed of vehicles that encountered AFAD was 23.2 mph, and the approaching speed of vehicles that encountered the human flagger was 27.4 mph. Approach speeds for vehicles that traveled through the AFAD were significantly lower than for the human flagger with a confidence level higher than 99.9 percent. Cohen's d indicated that the standardized mean of AFAD speed was 0.667 units of standard deviation lower than the mean of flagger.

Table 4.1.1.4 Speed of the Leading Vehicle at 250 ft. from the AFAD/Flagger

	Speed at 250 ft. (mph)
AFAD	23.23
Flagger	27.37
Confidence Level	> 99.9%*
Difference	-4.14
Cohen's d	-0.667
Effect Size	Medium

* indicates significance at 99% confidence level

MOE 2 measured the full stop location of vehicles that encountered STOP message/paddle. As shown in Table 4.1.1.5, the average full stop location of vehicles that encountered AFAD was 61.07 ft. behind the AFAD, and the average full stop location of vehicles that encountered human flagger was 49.64 ft. behind the flagger. The full stop location for AFAD was significantly farther away than the flagger with the confidence level being higher than 99.9 percent. Cohen’s d indicated the mean of AFAD full stop location was 0.436 units of standard deviation farther than flagger.

Table 4.1.1.5 Full Stop Location

	Full Stop Location (ft.)
AFAD	61.07
Flagger	49.64
Confidence Level	> 99.9%*
Difference	11.43
Cohen's d	0.436
Effect Size	Small

* indicates significance at 99% confidence level

MOE 3 measured the waiting time of the first vehicle in the queue, and MOE 7 measured the queue length in the stopped queue. MOEs 3 and 7 are shown in Table 4.1.1.6. These two MOEs were not related to safety but efficiency. Waiting time was defined as the time gap between vehicle restart and full stop. The waiting time for the AFAD was approximately 33 seconds less than the waiting time for the flagger. In some instances, the AFAD waiting time was increased because the AFAD showed "SLOW" on the CMS and the STOP paddle while vehicles were still clearing the work zone, thus requiring vehicles to wait for the opposing traffic to clear. An example of this situation is shown in Figure 4.1.1.14. One contributing factor to this situation was a synchronization delay between the STOP/SLOW paddle and the CMS. Although the synchronization issue has since been corrected, it is recommended that the AFAD operator ensures that all traffic has passed the end of the TMA (rather than the location of the AFAD operator) before switching the paddle and CMS from "STOP" to "SLOW".

Figure 4.1.1.14 Delay due to Opposing Traffic Not Clearing

Table 4.1.1.6 Waiting Time and Queue Length

	Waiting Time (s)	Queue Length (veh)
AFAD	72.25	1.70
Flagger	105.52	2.08
Confidence Level	99.8%*	99.4%*
Difference	-33.26	-0.39
Cohen's d	-0.389	-0.301
Effect Size	Small	Small

* indicates significance at 99% confidence level

MOE 4 measured the reaction time of the first vehicle in the queue. It was calculated as the time gap between the first appearance of SLOW message (AFAD) or paddle (flagger) and when the vehicle started to move again. As previously discussed, the reaction time based on the AFAD CMS was ultimately used instead of the AFAD paddle. As shown in Table 4.1.1.7, the average reaction time for the AFAD was 4.41 s, and for the flagger was 1.69 s. This result may be due to the differences in interpersonal communication with a person as opposed to interaction with a device. Another reason for the significant longer reaction time for drivers who encountered AFAD may be that some drivers were looking at their cellphones or were otherwise distracted, but the drivers that passed through the flagger may have been less distracted due to the presence of a live human flagger standing by the side. Also, as previously discussed, the lag between the CMS display and the paddle turning could also have been a factor. Cohen’s d (effect size) indicated that the mean reaction to AFAD was 2.921 units of standard deviation longer than reaction time to flagger. However, due to the synchronization time lag between AFAD paddle and AFAD CMS, no solid conclusion could be made on reaction time difference.

Table 4.1.1.7 Reaction Time (AFAD based on CMS, flagger based on paddle)

	Reaction Time (CMS) (s) Based on AFAD CMS	Reaction Time (CMS) (s) Based on AFAD Paddle
AFAD	4.41	0.412
Flagger	1.69	1.690
Confidence Level	> 99.9%*	> 99.9%*
Difference	2.72	-1.279
Cohen's d	2.921	-0.530
Effect Size	Large	Medium

* indicates significance at 99% confidence level

MOE 5 measured the noncompliance rate, which could be an indication of driver misunderstanding of the AFAD or flagger. Noncompliance refers to when a vehicle ignored the STOP sign, thus requiring the AFAD to honk its horn or the flagger to stop the vehicle using gestures. In some instances, the vehicle backed up to the proper position after the noncompliance. The noncompliance rate for AFAD was slightly lower than flagger, as shown in Table 4.1.1.8. However, the difference was not statistically significant. A previous MnDOT (2005) evaluation reported a noncompliance rate of 0.0096 (5/313). This is a similar low but non-negligible noncompliance rate.

Table 4.1.1.8 Noncompliance Rate

	Noncompliance Rate
AFAD	0.016 (3/193)
Flagger	0.019 (3/155)
Confidence Level	21.3%
Difference	-0.004
Cohen's d	-0.029
Effect Size	n/a

MOE 6 measured the approaching speed of the second vehicle in the queue. As shown in Table 4.1.1.9, the average speed of the second vehicle in the AFAD queue at 250 ft. was 20.6 mph, and in the flagger, queue was 23.1 mph. The difference was significant at the 99.5 percent confidence level. This result indicates that the second vehicle approached the AFAD at a lower speed than vehicles approaching the flagger.

Table 4.1.1.9 1st Following vehicle Speed at 250 ft.

	1st Following Vehicle Speed at 250 ft. (mph)
AFAD	20.63
Flagger	23.09
Confidence Level	99.5%*
Difference	-2.46
Cohen's d	-0.460
Effect Size	Small

* indicates significance at 99% confidence level

During the field collection process, unusual driving behavior was observed. Types of unusual driving behaviors include high approaching speed and extra-long reaction time. Two instances of high speeds at the flagger end were a pickup going 47 mph and an SUV going 55 mph (Figure 4.1.1.15). These two vehicles had approaching speeds which were much higher than the other vehicles since the average approaching speed was 27.4 mph. For long reaction times (Figure 4.1.1.16), one leading vehicle at the AFAD end had a reaction time of 20 seconds, while the average reaction time for AFAD was 4.41 s. After the CMS showed the SLOW sign, the leading vehicle did not realize the change of message on CMS, and the AFAD honked twice to get the vehicle’s attention.

Figure 4.1.1.15 Vehicle Approaching Flagger at High Speed


Figure 4.1.1.16 Vehicle Long Reaction Time to SLOW Indication on AFAD

Noncompliance at the south side of the work zone were less frequent than noncompliance at the north side of the work zone. One reason why the noncompliance rate at the north side was higher (Appendix B, Table B.2.2 and B.3.2) may be the difference in grades at the two ends. In the field study, the north end was at the top of a steep hill while the approach to the south end was more level. Some drivers may have wanted to know what was going on behind the stop control. At the south side, they could see more of the work zone as they approached, but at the north side, their view was more limited.

4.1.2 Field Survey

Survey Methodology

A driver intercept survey was administered for vehicles that traveled through the AFAD end of the work zone. Vehicles were stopped in the work zone after they went through the AFAD and given a short survey. There were two survey formats: hard copies with stamped envelopes and an index card with a link (including QR code) to an online version of the survey. In some cases, drivers were given a choice of which survey format they preferred. In other instances, to reduce vehicle delay, drivers were assigned a survey format based on the researcher's judgment of the survey format preference. For example, drivers who had their cell phones readily available or were texting on their phones were typically given the online version of the survey. The research team distributed 104 hard copies and 182 online links (Table 4.1.2.1). A total of 42 responses were received, and the response rate was 14.7 percent. This response rate is relatively low but is similar to some of the mail surveys discussed in Hager et al. (2003).

Table 4.1.2.1 Survey Numbers

Survey	Hard Copy	Online	Total
Sent Out	104	182	286
Response Received	30	12	42

The survey consisted of four parts. Parts 1 and 2 asked questions about drivers' understanding of the AFAD signage and human flagger gesture, their perceptions regarding the effectiveness of the two different stop controls, their opinion regarding whether the CMS was helpful, and any additional comments. Part 3 asked for their preference between the AFAD and flagger. Part 4 asked for their demographic information and regular vehicle type. The complete field survey is attached in Appendix C.

Survey responses included two types: hard copies and online. To ensure consistency in survey data processing, hard copy entries were entered into the online survey system. Results were extracted directly from the online survey system.

Survey Results

Two multiple choice questions involved the meaning of the AFAD signage and human flagger gesture, respectively. Among the 42 respondents, all of them understood the AFAD meaning correctly, but two of them chose the wrong answer for the meaning of the flagger gesture. The results implied that the AFAD was more understandable than the flagger.

The survey responses indicate that most of the respondents thought both AFAD and flagger were effective. Although 88.1 percent of respondents thought AFAD was effective or very effective and 92.86 percent of respondents thought flagger was effective or very effective, the proportion of respondents who thought AFAD was very effective was more than the proportion who thought that the flagger was very effective. However, there were more respondents who thought that AFAD was ineffective or very ineffective, as shown in Table 4.1.2.2. This result may due to the novelty of the AFAD as these drivers had not previously encountered the AFAD. Some drivers may have preferred the interpersonal communication with the flagger.

Table 4.1.2.2 Survey Responses Regarding Effectiveness

Effectiveness	STOP/SLOW AFAD			Flagger		
	Count	Percentage		Count	Percentage	
Very Effective	28	66.67%	88.10%	8	19.05%	92.86%
Effective	9	21.43%		31	73.81%	
Neutral	1	2.38%	2.38%	1	2.38%	2.38%
Ineffective	1	2.38%	9.52%	1	2.38%	4.76%
Very Ineffective	3	7.14%		1	2.38%	
Total	42	100.00%		42	100.00%	

Respondents were asked about the reasons for their effectiveness ratings for the AFAD and flagger. Five factors were provided as possible answers: clarity, visibility, safety, efficiency, and other. Among the four factors, visibility ranked number one, in both AFAD and flagger situations as shown in Table 4.1.2.3. Clarity and safety were also both considered as important reasons for the effectiveness ratings.

Table 4.1.2.3 Reason of Effectiveness Rating

Factor	Count		
	AFAD	Flagger	Total
Clarity	21	31	52
Visibility	23	36	59
Safety	20	30	50
Efficiency	13	20	33
Other	5	5	10

As shown in Table 4.1.2.4, 90.48 percent of the respondents thought that the CMS was helpful, with 57.14 percent of the respondents strongly in agreement. Only one respondent (2.38 percent) disagreed or strongly disagreed that the CMS was helpful. Most of the respondents thought that the CMS improved the visibility of stop control and could help them to understand signage. One respondent felt that the CMS was redundant and unnecessary since the STOP/SLOW paddle was present and was informative enough.

Table 4.1.2.4 Summary of Responses to Survey Question Regarding Helpfulness of CMS

CMS helpfulness	Count	Percentage	
Strongly Agree	24	57.14%	90.48%
Agree	14	33.33%	
Neutral	3	7.14%	7.14%
Disagree	0	0.00%	2.38%
Strongly Disagree	1	2.38%	
Total	42	100.00%	

The survey asked if the drivers had encountered the two types of stop controls before. Although the respondents had just driven through the AFAD, less than half of them responded that they had encountered an AFAD before, while all of them had previously encountered a flagger (Table 4.1.2.5).

Table 4.1.2.5 Summary of Responses to Question about Previous Experience with AFAD and Flagger

Encountered Before?	AFAD		Flagger	
	Count	Percentage	Count	Percentage
Yes	19	45.24%	41	100%
No	23	54.76%	0	0%
Total	42	100.00%	41	100%

When drivers were asked for their preference between AFAD and flagger, no respondents preferred the flagger much more than AFAD, and only 12.2 percent of the respondents preferred the flagger more. Although the percentage of respondents who thought that the flagger was effective or very effective was higher than the percentage who thought that the AFAD was effective or very effective, respondents preferred the AFAD more than the flagger. As shown in Table 4.1.2.6, 53.66 percent respondents preferred the AFAD much more than flagger, and 24.39 percent preferred the AFAD more than flagger.

Table 4.1.2.6 Respondents' Preference for AFAD or Flagger

Preference	Count	Percentage	
AFAD much more	22	53.66%	78.05%
AFAD more	10	24.39%	
Neutral	4	9.76%	9.76%
Flagger more	5	12.20%	12.20%
Flagger much more	0	0.00%	
Total	41	100.00%	

Demographic information was collected, and the results are shown in Tables 4.1.2.7 and 4.1.2.8. Among the survey respondents, gender distributions were even, with the number of female drivers slightly less than the number of male drivers. Older drivers were more prevalent than younger drivers, and over 64 percent of the respondents were over 55 years old. The field work was performed in a rural area, and 83.33 percent of respondents were rural residents. Most of the respondents drove passenger cars as their regular vehicle type. Different responses by age, gender, and residency are attached in Appendix D.

Table 4.1.2.7 Demographic Information

Gender		**Age**				
Male	Female	16-25	26-40	41-55	56-70	71-95
22	19	1	5	8	14	13
52.38%	45.24%	2.38%	11.90%	19.05%	33.33%	30.95%

Table 4.1.2.8 Residency and Vehicle Information

Residency		**Regular Vehicle Type**	
Urban	Rural	Passenger car	Other
3	35	37	4
7.14%	83.33%	88.10%	9.52%

Respondents provided written comments on the advantages and disadvantages of AFAD. They thought the advantages of AFAD included increased visibility, multi-functionality, adaptability to weather conditions, and enhanced safety, as a human flagger means a worker is standing near traffic. Some concerns raised by some respondents about the AFAD included:

- Sun glare reduced visibility

- Potential confusion in case of its malfunction
- AFAD may not be respected as well as a live human flagger
- It may be easier to communicate with human flaggers than the AFAD

Some additional comments include:

- The higher cost of AFAD was worthwhile due to its benefits
- A warning noise for violations would help to alert both drivers and workers in the work zone
- Advanced signage for TMA instructions would be beneficial
- Malfunction of AFAD

4.1.3 Simulator Study

After Phase One, field study of the MoDOT AFAD and human flagger, was completed, Phase Two, simulator study, was conducted. The simulator was utilized to examine various details of the AFAD design in a cost-efficient manner since AFAD variations were implemented in a virtual world. This simulator study explored three different AFAD configurations, including one without the use of a CMS. The simulator offered a highly controlled environment, thus limiting extraneous causal factors. The simulator also provided safe experiment conditions to allow for different options to be tested.

Simulator Study Methodology

ZouSim, the University of Missouri’s driving simulator, was used for conducting the AFAD simulator study. ZouSim is a medium-fidelity simulator built around the half-cab of a sedan. Even though ZouSim has a wide range of graphical display capabilities, including virtual reality, augmented reality, and stereoscopic 3D, the triple 120-inch screen was chosen as the most appropriate display for the AFAD study. This display setup provided a 180-degree field-of-view which offered an excellent view of the approaching work zone and the relevant peripheral clues for regulating driving speed. Figure 4.1.3.1 shows the ZouSim setup for the AFAD experiment. The primary virtual camera was the forward windshield view. Three additional virtual cameras presented the left, right and rear-view mirrors perspectives. The active instrumentation in the vehicle includes a force-feedback steering wheel, brake and acceleration pedals, turn signals, and an engine vibration generator.

Figure 4.1.3.1 AFAD Experiment Using ZouSim

<u>Scenarios</u>

The study simulated work zones on a rural highway in Missouri with a speed limit of 55 mph. The entire highway was designed without any horizontal or vertical curves in order to eliminate the influence of terrain. The road was created according to AASHTO Green Book standards (AASHTO 2013). Surfaces were textured and/or painted with the appropriate striping and markings that conform to the MUTCD (FHWA 2009). The work zone plan on the highway is shown in Figure 4.1.3.2. In the work zone, one lane is closed for road work while traffic movements in the open lane are controlled by a pair of flaggers or AFADs. The work zone configuration, including signage, distance between signs, and location of flagger/AFAD, followed the requirements of MUTCD (FHWA 2009) and MoDOT Engineering Policy Guide (MoDOT 2009).

As shown in Figure 4.1.3.2, the following four flagging methods, one flagger and three AFAD designs, were tested: (1) human flagger, (2) MoDOT AFAD, (3) variation on method (2) with an alternative sign, and (4) variation on method (2) with the CMS turned off. There was no other traffic in the direction that the participant is traveling, so the participant was always the leading vehicle. When arriving at a work zone, participants always encountered the STOP interval and had to wait for the opposite traffic to pass first before given the SLOW sign.

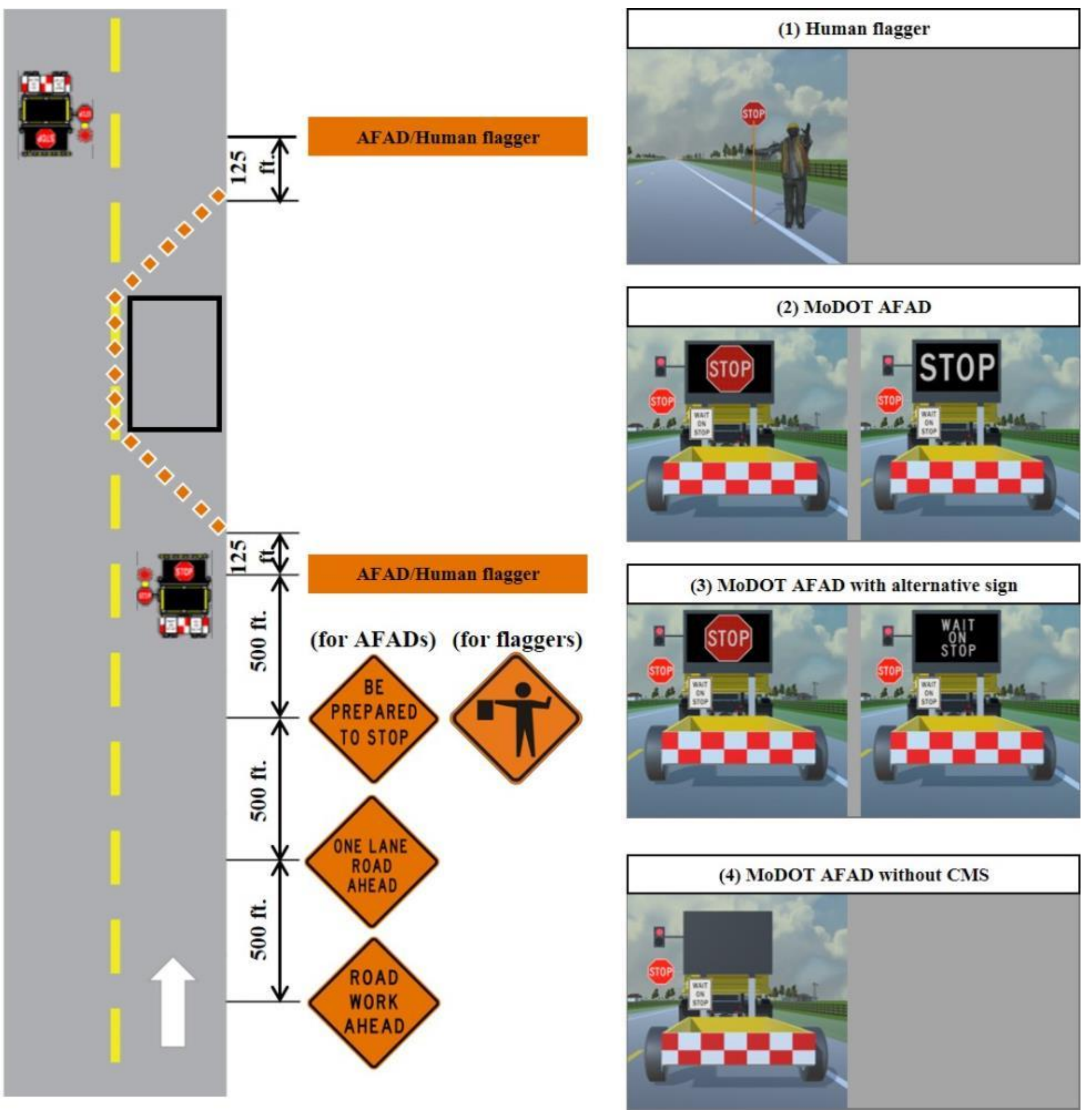


Figure 4.1.3.2 Work Zone Plan and Traffic Control Methods in the Simulator Study

The human flagger was the baseline for this experiment. A human flagger manually operated a STOP/SLOW paddle. The size of paddle was an 18-inch hexagon conforming to the MUTCD (FHWA 2009). There were two faces of the paddle: one face showed STOP in a red hexagon and the other one showed SLOW in an orange rhombus. During the STOP phase, the human flagger stopped traffic according to the MUTCD flagger guidance (FHWA 2009) while the flagger waved his arm to guide traffic during the SLOW phase as shown in Figure 4.1.3.3.

Figure 4.1.3.3 Human Flagger Configuration

The AFAD was the simulation of the MoDOT AFAD deployed in field tests. The AFAD was mounted on a TMA with a beacon, a STOP/SLOW paddle, and a CMS. There were two phases: STOP and SLOW. During the STOP phase, the beacon was static red, the paddle showed the STOP face, and the message on the CMS alternated between a STOP sign and the word “STOP”, as shown in Figure 4.1.3.4. During the SLOW phase, the message on the CMS alternated between a SLOW sign and the phrase “GO ON SLOW.”

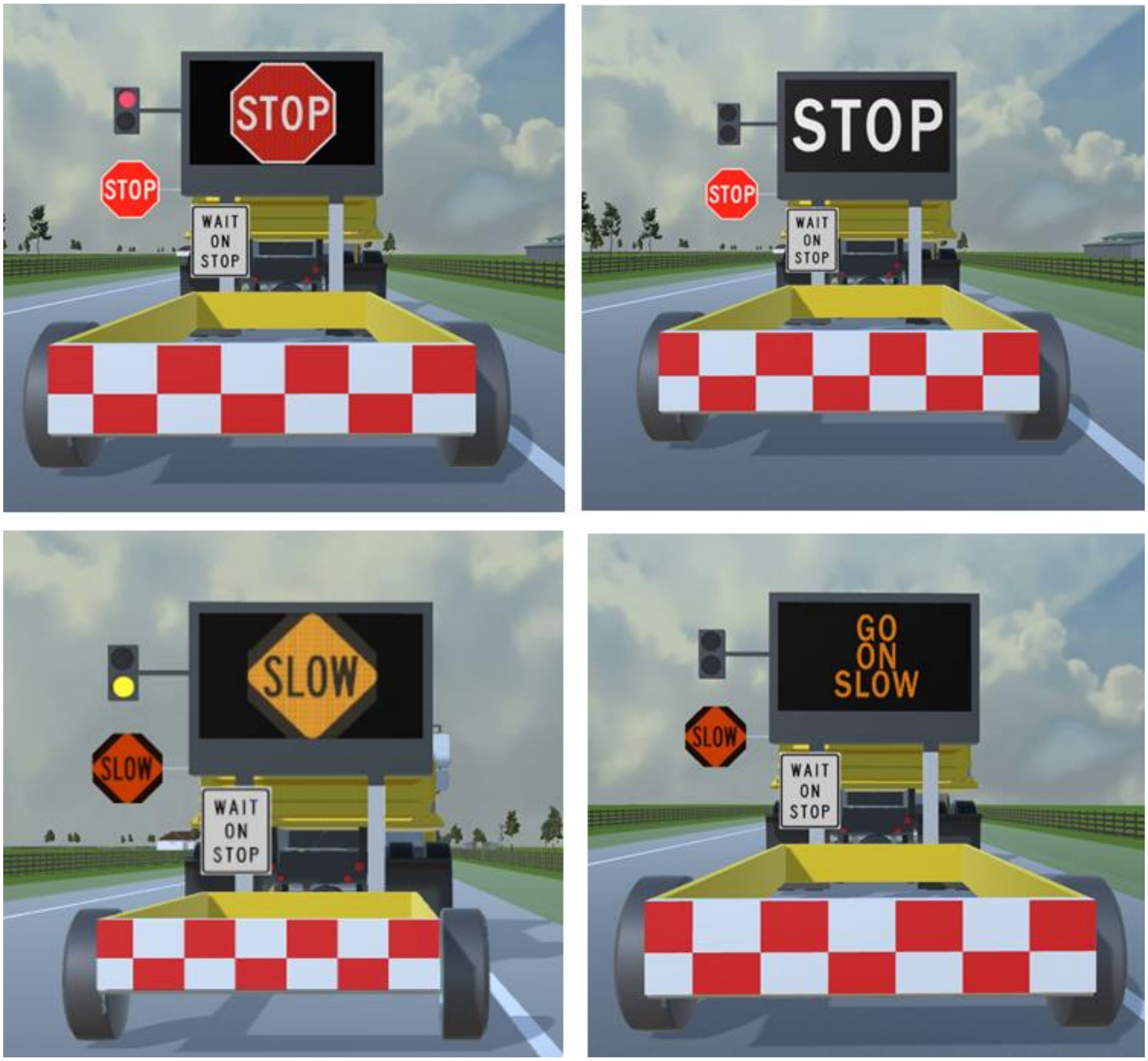


Figure 4.1.3.4 MoDOT AFAD Configuration

The AFAD with the alternative sign was similar to MoDOT AFAD. The only variation was the STOP phase with "WAIT ON STOP" being displayed on the CMS instead of the word "STOP", after the STOP sign was shown on the CMS. The configuration of the AFAD with the alternative sign is shown in Figure 4.1.3.5.

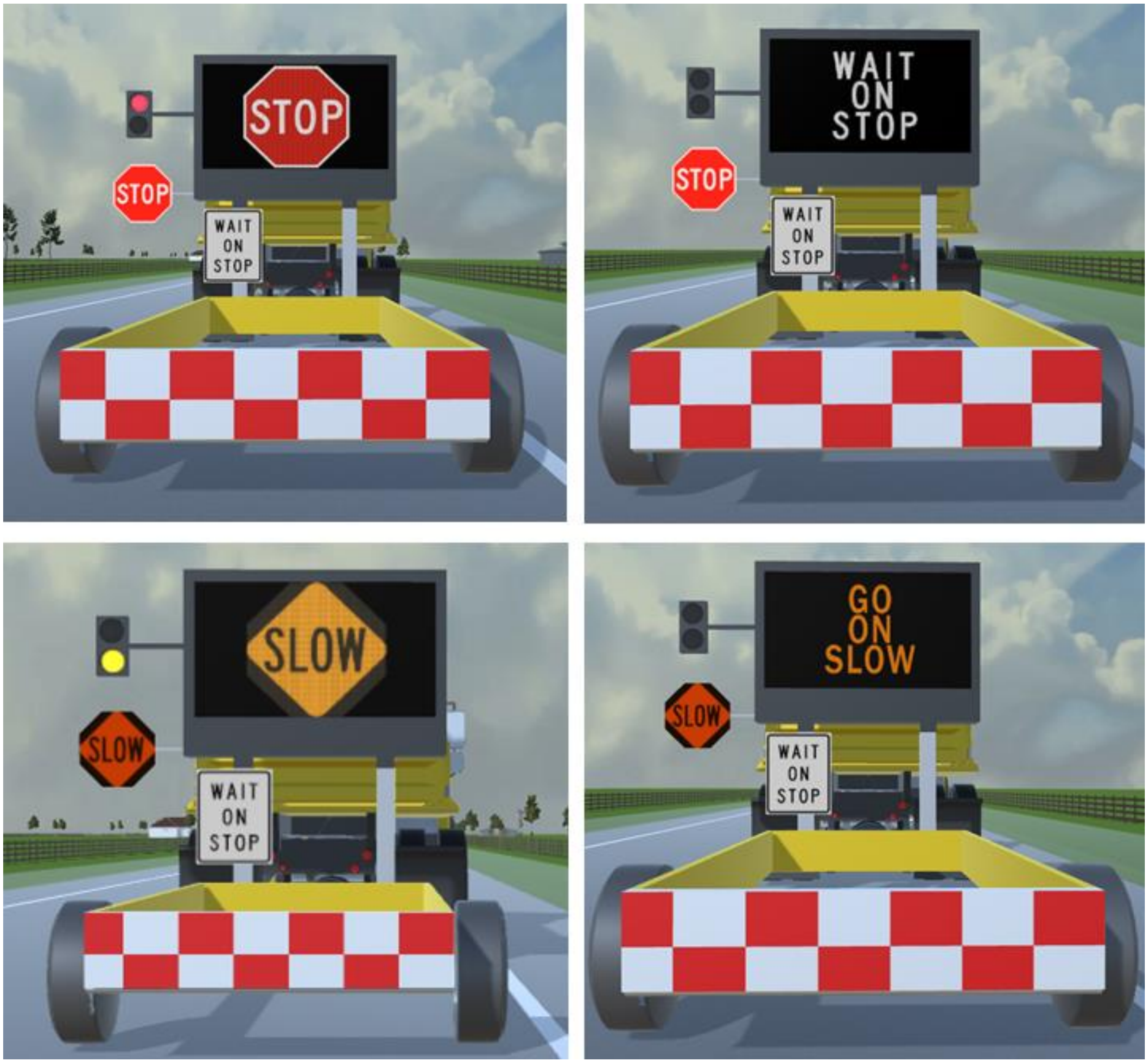


Figure 4.1.3.5 AFAD with Alternative Sign Configuration

The AFAD without a CMS was a simplified version of the MoDOT AFAD with the CMS turned off, as shown in Figure 4.1.3.6. The CMS was not removed for AFAD without CMS as per MoDOT request in order to replicate an AFAD with a CMS-style arrow board. Thus, a MoDOT AFAD would not have the CMS space unoccupied without an arrow board. This arrow board was not turned on for the simulator experiment as per MoDOT request.

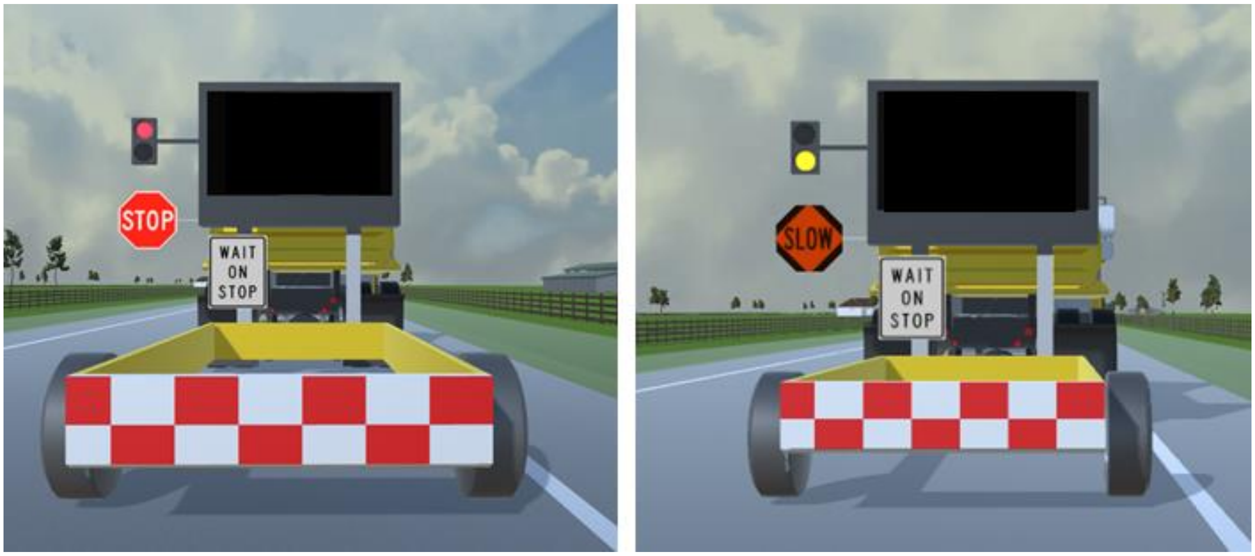


Figure 4.1.3.6 AFAD without CMS Configuration

In order to minimize sequence bias, with four different flagging methods, 24 different test orders were generated, and each participant was randomly assigned to one of 24 test orders to minimize sequence bias effects. Each of the four flagging methods was tested twice to obtain more data. The waiting time was varied, either 30 seconds or 40 seconds, to prevent participants from recognizing a pattern in waiting times.

Simulator Trials and MOEs

The study protocols and measurement tools were evaluated and approved by the campus institutional review board, and a standard hosting script was used. First, a participant's informed consent was obtained after being introduced to the simulator and the experiment purpose. Then, the participant drove a free-driving warm-up scenario to become familiar with the specific vehicle. Once the participant was comfortable, the actual work zone scenarios were initiated. The participant was asked to drive along a rural two-lane, two-way highway to arrive at a clinic. On the way, the participant encountered eight work zones involving the four flagging methods.

After the simulator data was collected, the research team reviewed the videos, and conducted the data reduction process. Five Measure of Effectiveness (MOEs) were defined for data reduction as described below.

- MOE 1 - Approach Speed (mph) is the speed when a vehicle is 250 feet from the flagger/AFAD. This is the same distance at which field speed data was collected. At this location, drivers can see the AFAD/flagger clearly. A lower approach speed is desirable. The approach speed was echoed on the screen as shown in Figure 4.1.3.7. It also shows that the vehicle was at 250 ft. from the flagger/AFAD.

Figure 4.1.3.7 MOE 1 Example: Approach Speed (mph)

- MOE 2 - Full Stop Distance (feet) is the distance to the human flagger or AFAD when the vehicle fully stops. A larger distance means more separation from the flagger/AFAD. This is the same MOE recorded in the field. MOE 2 was echoed on the screen as shown in Figure 4.1.3.8.

Figure 4.1.3.8 MOE 2 Example: Full Stop Distance (feet)

- MOE 3 - Reaction Time (seconds) is the time between when "SLOW" sign is shown and when the vehicle restarts. A faster reaction may indicate that the driver has a better understanding of when to go. An example of MOE 3 is shown in Figure 4.1.3.9.

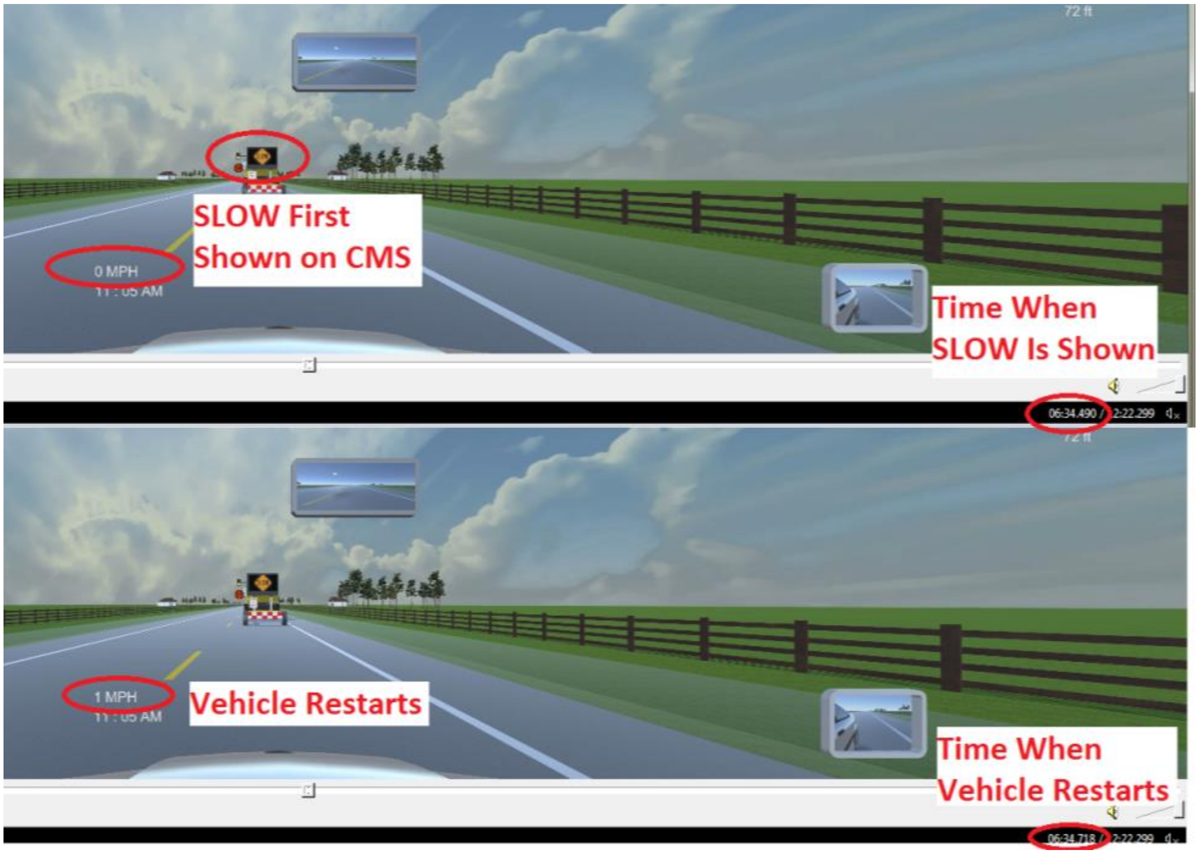


Figure 4.1.3.9 MOE 3 Example: Reaction Time (seconds)

- MOE 4 - Noncompliance Rate is the percentage of drivers not obeying the flagging instructions as indicated by a driver not stopping or stopping after the location of the flagger/AFAD. A lower noncompliance rate is desired. An example of a driver trying to bypass the AFAD is shown in Figure 4.1.3.10.

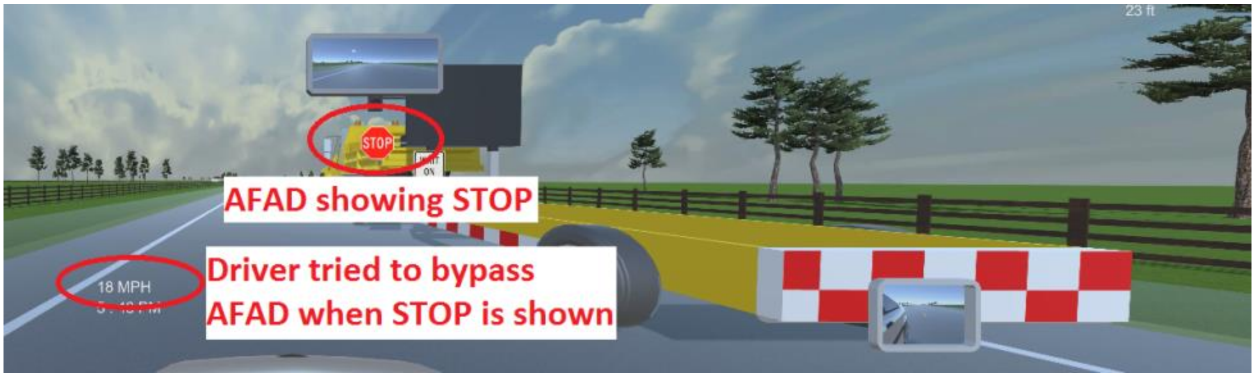


Figure 4.1.3.10 MOE 4 Example: Noncompliance Rate

- MOE 5 - First Brake Location (feet) is the location at which the driver first pressed the decelerator pedal in reaction to the flagger/AFAD. This location could indicate where a driver recognized and understood the flagging instruction. MOE 5 is illustrated in the screen output as shown in Figure 4.1.3.11.

Figure 4.1.3.11 MOE 5 Example: First Brake Location (feet)

<u>Simulator Study Results</u>

The video of each participant trial was recorded and all of the MOEs were extracted from videos and presented. The MOEs were purposely extracted from video instead of an automated data file since visual confirmation is often helpful for identifying data issues. To compare MOEs between the flagger and AFAD alternatives, differences between them were calculated. Statistical analysis was performed to calculate significance, confidence level, and effect size. Confidence level higher than 95 percent was regarded as significant in this study. Effect size was presented as Cohen's d value, and significant difference was defined to be small, medium, and large (Cohen 1977), as effect size between 0.2 and 0.5 was small, effect size between 0.5 and 0.8 was medium, and effect size > 0.8 was large.

To clarify comparisons and provide guidance to decision makers, the results were divided into two parts. The first part is the comparison between human flagger and MoDOT AFAD, with the flagger as the baseline. The second part is the comparison among the three AFAD scenarios, with the MoDOT AFAD as the baseline. The test result in its totality is attached in Appendix E.

Simulator Results: MoDOT AFAD vs. Human Flagger

MOE 1 measured the speed of the vehicle at 250 ft. from the MoDOT AFAD/flagger. As shown in Table 4.1.3.1, the average speed of vehicles approaching the MoDOT AFAD was 26.3 mph, and the speed of vehicles approaching the human flagger was 34.8 mph. Approach speeds for vehicles that traveled through the MoDOT AFAD were significantly lower than for the human flagger with a confidence level higher than 99.9 percent. Cohen's d indicated that the standardized mean of AFAD speed was 0.66 units of standard deviation lower than the mean for flagger, and this difference was considered a medium effect size. The results show that MoDOT AFAD was more visible and safer than a human flagger.

Table 4.1.3.1 Approach Speed MoDOT, AFAD vs. Human Flagger

	Approach Speed (mph)
MoDOT AFAD	26.34
Flagger	34.79
Difference	-8.44
Confidence Level	> 99.9%*
Cohen’s d	-0.66
Effect Size	Medium

* indicates significance at 99% confidence level

MOE 2 measured the full stop location of vehicles that encountered STOP message/paddle. As shown in Table 4.1.3.2, the average full stop location of vehicles that encountered AFAD was 97.55 ft. in front of the AFAD, and the average full stop location of vehicles that encountered human flagger was 53.09 ft. in front of the flagger. The full stop location for AFAD was significantly farther away than the flagger with the confidence level being higher than 99.9 percent. Cohen's d indicated the mean of AFAD full stop location was 1.02 units of standard deviation farther than flagger, which was a large effect size difference. Therefore, MoDOT AFAD improves work zone safety by encouraging drivers to stop farther away.

Table 4.1.3.2 Full Stop Distance, MoDOT AFAD vs. Human Flagger

	Full Stop Distance (ft.)
MoDOT AFAD	97.55
Flagger	53.09
Confidence Level	> 99.9%*
Difference	44.46
Cohen's d	1.02
Effect Size	Large

* indicates significance at 99% confidence level

MOE 3 measured the reaction time of the driver. It was calculated as the time gap between the first appearance of the SLOW paddle for both flagger and AFAD, and when the vehicle started to move again. As shown in Table 4.1.3.3, the average reaction time for the AFAD was 1.93 s, and for the flagger was 2.05 s. The difference was not significant. It indicated that drivers reacted similarly when the SLOW indication was presented either by the MoDOT AFAD or the human flagger.

Table 4.1.3.3 Reaction Time, MoDOT AFAD vs. Human Flagger

	Reaction Time (s)
MoDOT AFAD	1.93
Flagger	2.05
Confidence Level	70.3%
Difference	-0.12
Cohen's d	-0.03
Effect Size	n/a

MOE 4 measured the noncompliance rate, which is an indication of a driver's misunderstanding of the AFAD or flagger. Noncompliance refers to when STOP is indicated, and the driver bypassed the AFAD or flagger without fully stopping. Among the 64 iterations from 32 participants (each participant went through each scenario twice), no noncompliance events happened with the MoDOT AFAD, and nine noncompliance occurred for the human flagger, as shown in Table 4.1.3.4. The noncompliance rate indicated that the MoDOT AFAD was more understandable and communicated information to drivers more effectively.

Table 4.1.3.4 Noncompliance Rate, MoDOT AFAD vs. Human Flagger

	Noncompliance Rate
MoDOT AFAD	0.00 (0/64)
Flagger	0.14 (9/64)
Confidence Level	99.8%*
Difference	-0.14
Cohen's d	n/a
Effect Size	n/a

* indicates significance at 99% confidence level

MOE 5 measured the first brake location where the driver first hit the brake pedal in reaction to the flagger/AFAD. This measure could indicate where a driver first recognized and understood the flagging instruction. The results in Table 4.1.3.5 showed that the first brake location for MoDOT AFAD was significantly farther away, and the difference was a medium effect size. This measure indicated that the MoDOT AFAD was more visible than the human flagger, and the MoDOT AFAD clearly conveyed the STOP message to drivers.

Table 4.1.3.5 First Brake Location, MoDOT AFAD vs. Human Flagger

	First Brake Location (ft.)
MoDOT AFAD	332.19
Flagger	274.02
Confidence Level	99.5%*
Difference	58.17
Cohen's d	0.51 (Medium)

* indicates significance at 99% confidence level

Overall, the comparison between the MoDOT AFAD and human flagger showed that MoDOT AFAD performed significantly better than human flagger with respect to approach speed, full stop distance, noncompliance rate, and first brake location. The results imply that the

MoDOT AFAD may improve work zone safety and can be a valid and effective replacement for the human flagger.

Simulator Results: MoDOT AFAD vs. other AFADs

MOE 1 measured the speed of the vehicle at 250 ft. from AFADs. As shown in Table 4.1.3.6, the average speeds of vehicles approaching AFADs were similar for all three AFAD configurations, and the differences were not significant.

Table 4.1.3.6 Approach Speed, MoDOT AFAD vs. other AFADs

	MOE 1 Approach Speed (mph)			
	Mean	Diff	Cohen's d	Confidence Level
MoDOT AFAD	26.34	Baseline		
AFAD with alternative sign	25.98	-0.36	0.03	28.2%
AFAD without CMS	26.87	0.53	0.05	33.7%

MOE 2 measured the full stop location of vehicles from AFADs, as shown in Table 4.1.3.7. The average full stop distances for the MoDOT AFAD and the AFAD with alternative sign were similar. However, vehicles that encountered the AFAD without CMS stopped 23.35 ft. closer to the AFAD than vehicles that encountered the MoDOT AFAD, which was a medium effect size. This result indicates that the CMS may be needed to better communicate STOP instructions to drivers.

Table 4.1.3.7 Full Stop Location, MoDOT AFAD vs. other AFADs

	MOE 2 Full Stop Distance (feet)			
	Mean	Diff	Cohen's d	Confidence Level
MoDOT AFAD	97.55	Baseline		
AFAD with alternative sign	90.67	-6.88	-0.14	89.7%
AFAD without CMS	74.20	-23.35	-0.58 (Medium)	> 99.9%*

* indicates significance at 99% confidence level

MOE 3 measured the reaction time of the driver. Drivers reacted to the AFAD with alternative sign and the AFAD without CMS faster than the MoDOT AFAD as shown in Table 4.1.3.8. The difference between MoDOT AFAD and the AFAD with alternative sign was not significant. However, the reaction time to the AFAD without CMS was significantly shorter than the MoDOT AFAD. This result indicates that the presence of the CMS may have required more processing time for drivers after the SLOW indication was presented.

Table 4.1.3.8 Reaction Time, MoDOT AFAD vs. other AFADs

	MOE 3 Reaction Time (seconds)			
	Mean	Diff	Cohen's d	Confidence Level
MoDOT AFAD	1.93	Baseline		
AFAD with alternative sign	1.60	-0.32	-0.13	82.8%
AFAD without CMS	1.23	-0.70	-0.42 (Small)	98.8%**

** indicates significance at 95% confidence level

MOE 4 measured the noncompliance rate, which is an indication of a driver's understanding of AFADs. Among the 64 iterations from 32 participants (each participant went through each scenario twice), no noncompliance occurred when drivers encountered the MoDOT AFAD or the AFAD with alternative sign, and three noncompliance occurred for the AFAD without CMS, as shown in Table 4.1.3.9. Differences in noncompliance rates were not significant among all three AFAD configurations.

Table 4.1.3.9 Noncompliance Rate, MoDOT AFAD vs. other AFADs

	MOE 4 Noncompliance Rate			
	Mean	Diff	Cohen's d	Confidence Level
MoDOT AFAD	0.00	Baseline		
AFAD with alternative sign	0.00	0.00	n/a	n/a
AFAD without CMS	0.05 (3/64)	0.05	n/a	91.7%

MOE 5 measured the first brake location where the driver first applied the brake pedal in reaction to AFADs, which indicates where a driver recognized and understood the flagging instruction. The results in Table 4.1.3.10 show that the first brake location for all three AFAD configurations were similar and the differences were not significant.

Table 4.1.3.10 First Brake Location, MoDOT AFAD vs. other AFADs

	MOE 5 First Brake Location (feet)			
	Mean	Diff	Cohen's d	Confidence Level
MoDOT AFAD	332.19	Baseline		
AFAD with alternative sign	334.95	2.77	-0.03	13.9%
AFAD without CMS	320.30	-11.89	-0.11	46.6%

Overall, the comparison of the three AFAD configurations showed that the performance of AFAD with the alternative CMS message was similar to the performance of the MoDOT AFAD, and no significant differences occurred due to the replacement of the "STOP" message with "WAIT ON STOP". The AFAD without CMS resulted in a significant closer full stop distance than the MoDOT AFAD which may indicate that it was less visible and drivers approached closer to see the STOP/SLOW paddle more clearly. In addition, the reaction time for

AFAD without CMS was significantly shorter than MoDOT AFAD which may indicate that the CMS required additional processing time for drivers when changing to the SLOW indication. Although the difference in noncompliance rate was not significant, the AFAD without CMS experienced three noncompliance while the MoDOT AFAD had no noncompliance. The results indicate that the AFAD without CMS did not communicate information to drivers as effectively as the MoDOT AFAD.

4.1.4 Post-simulator Survey

Survey Methodology

A post-experiment survey was administered to obtain stated preferences and qualitative feedback from the simulator study participants. The 15-question survey consisted of four parts. Part 1 asked questions about participants' understanding of the AFAD signage and human flagger gestures. Part 2 asked participants to indicate their preferences and rate the clarity, visibility, safety, and efficiency properties of designs, as well as the necessity of the CMS. Part 3 of the survey asked about the fidelity of the driving simulator while demographic information of the participants was collected in Part 4. A 16-question simulator sickness questionnaire (SSQ) was provided at the end, which is widely used in diagnosing the simulator sickness severity of participants (Kennedy et al. 1993). The complete simulator survey is shown in Appendix F.

Participant recruitment was open to the public. During the recruitment process, an effort was made to align to the demographic distribution of the field study to the extent possible. However, due to differences in demographics between the locations of the field and simulator studies, the demographics were not fully matched. Balance in the distribution of age and gender also was sought. The total number of participants was 32, and demographic information is shown in Tables 4.1.4.1 and 4.1.4.2. Approximately 78 percent of participants were between the ages of 18 and 40, and 62.5 percent of the participants were male. Because the experiment was conducted in an urban area, over 96 percent of participants were from an urban area, and only one participant was a rural resident. Over 90 percent drove passenger cars as their regular vehicle, and less than 10 percent of them drove other vehicles, such as a truck or bus.

Table 4.1.4.1 Age and Gender Distribution

	Age					Gender	
	18-25	26-40	41-55	56-70	71-95	Male	Female
Count	12	13	4	2	1	20	12
Percentage	37.50%	40.63%	12.50%	6.25%	3.13%	62.50%	37.50%

Table 4.1.4.2 Residency and Regular Vehicle Type

	Residency		Regular Vehicle Type	
	Urban	Rural	Passenger Car	Other
Count	31	1	29	3
Percentage	96.88%	3.13%	90.63%	9.38%

Survey Question Results

Part 1 of the survey asked for participants' understanding of the flagger and three AFAD configurations. From the results of Question 1 to 4 (Table 4.1.4.3), the human flagger had the best comprehension with a rate of understanding of 94 percent, followed by the AFAD with alternative sign (91 percent). The MoDOT AFAD had an understanding rate of 84 percent, and the AFAD without CMS was understood by 81 percent of participants. Both participants who misunderstood the human flagger regarded it as a regular STOP sign, and all the participants who misunderstood the AFADs thought they were traffic lights.

Table 4.1.4.3 Drivers' Understanding of AFADs/Flagger

Drivers' Understanding	Sample Size	Answered Correctly	Correct Rate
Flagger	32	30	93.75%
MoDOT AFAD	32	27	84.38%
AFAD with Alternative Sign	32	29	90.63%
AFAD without CMS	32	26	81.25%

Part 2 of the survey asked participants to rank and rate the AFADs and flagger and assess the necessity of the CMS. In Question 5, participants were asked to rank their preference of the four different setups with “1” representing the most preferred, and “4” representing the least preferred. The results from this question are summarized in Table 4.1.4.4. The MoDOT AFAD had the highest average ranking of 1.53 while the AFAD without CMS had the lowest average ranking of 3.47. These results demonstrate that the MoDOT AFAD was the most preferred setup followed by the AFAD with alternative sign and the human flagger. The AFAD without CMS was the least preferred setup.

Table 4.1.4.4 Preference Ranking

Preference	Average Score	Ranking
Flagger	3.07	3
MoDOT AFAD	1.53	1
AFAD with Alternative CMS Message	1.90	2
AFAD without CMS	3.47	4

To better understand how participants ranked their preference and see if there were any correlations between the rankings, one-way ANOVA tests were conducted. The plots of these results are shown in Figures 4.1.4.1, 4.1.4.2, and 4.1.4.3. The plots show the distribution of scores with a line connecting the mean scores. From Figure 4.1.4.1, it can be seen that the majority of participants who preferred the MoDOT AFAD the most, preferred the human flagger the least. The preferences for the MoDOT AFAD and the AFAD with alternative sign were similar, and their rankings were close to each other as shown in Figure 4.1.4.2. The AFAD without CMS was not preferred by participants who were in favor of the MoDOT AFAD, as shown in Figure 4.1.4.3.

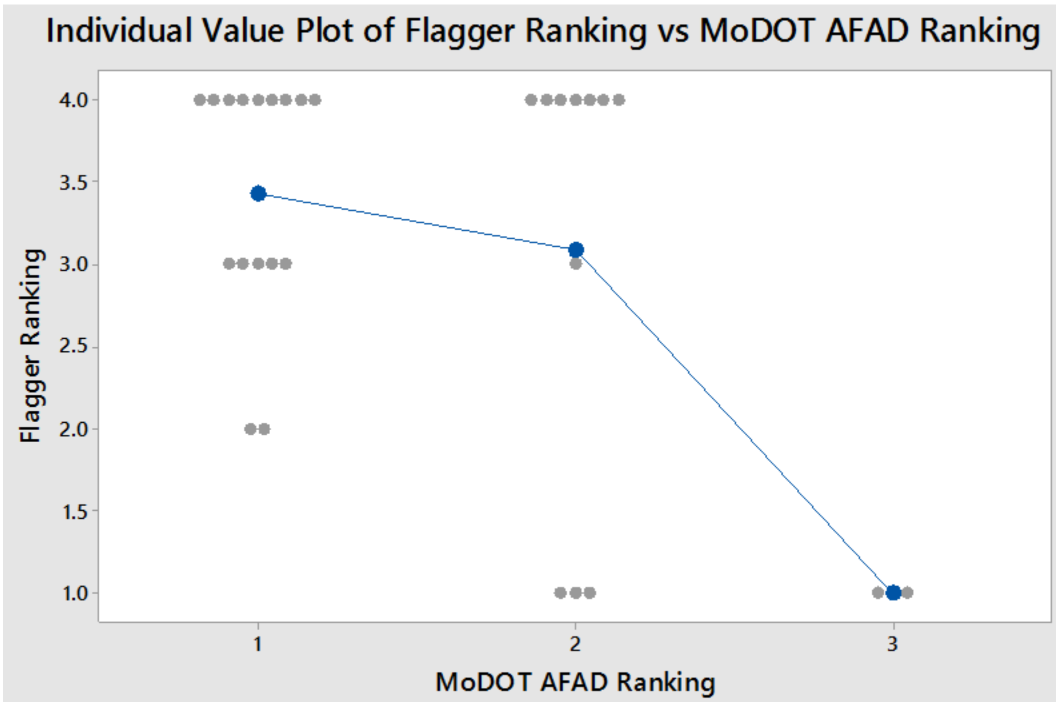


Figure 4.1.4.1 Ranking: MoDOT AFAD vs. Flagger

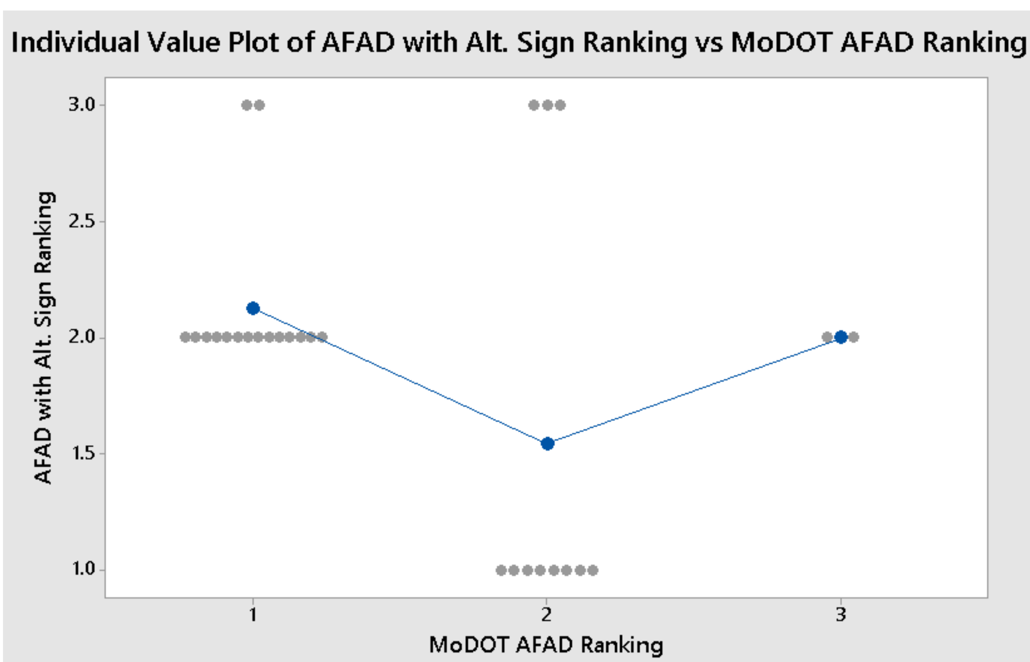


Figure 4.1.4.2 Ranking: MoDOT AFAD vs. AFAD with Alternative Sign

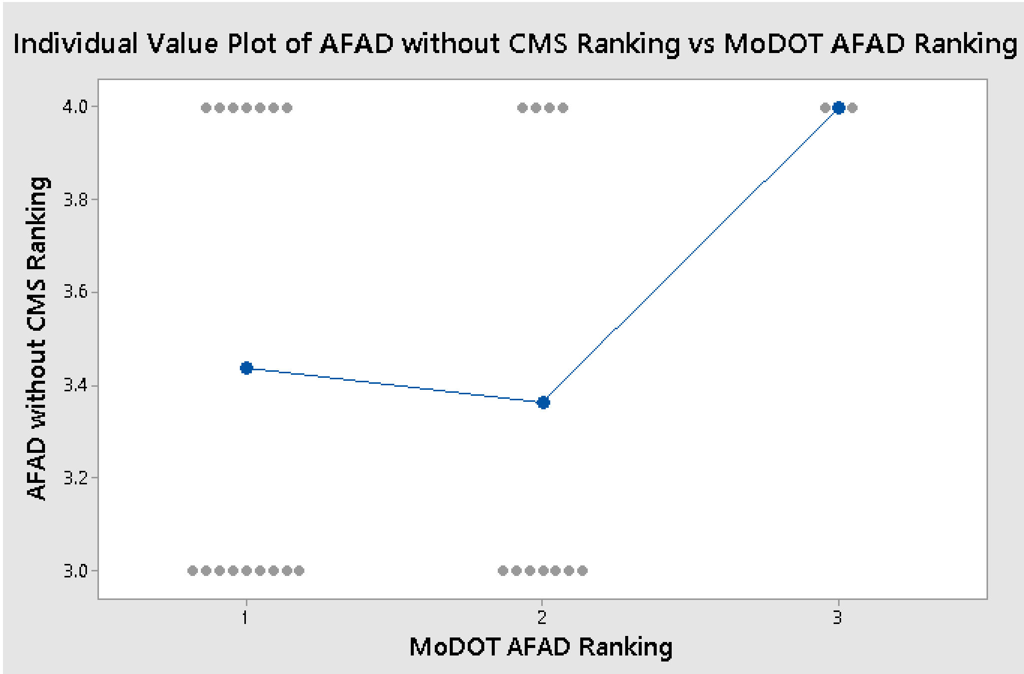


Figure 4.1.4.3 Ranking: MoDOT AFAD vs. AFAD without CMS

The survey also asked participants to rate different attributes of the human flagger and AFAD configurations in Question 6. These attributes included clarity, visibility, safety, and efficiency. The average score across the four attributes for each setup was consistent, as shown in Table 4.1.4.5. The MoDOT AFAD scored the highest in all four attributes, and the AFAD with alternative sign scored the second highest in all categories. The AFAD without CMS scored the lowest in clarity but was higher than the human flagger for the other three attributes. The human flagger had the lowest scores in visibility, safety, and efficiency. Since these ratings were subjective discrete ordinal data, a nonparametric Mann-Whitney test was conducted to assess the statistical significance of the ranked data. The Mann-Whitney test is used to assess ordinal data that is not normally distributed by calculating average score differences and determining if difference between the data sets is significant (De Winter and Dodou 2010). The confidence level for the range of differences was set to be 95 percent in this study. The results from this test showed that the AFAD without CMS and the human flagger had significantly lower scores than the MoDOT AFAD for all four attributes, and the AFAD with alternative CMS message scored significantly lower in visibility and safety than the MoDOT AFAD. One-way ANOVA results are attached in Appendix G.

Table 4.1.4.5 Ratings for Clarity, Visibility, Safety, and Efficiency

	Clarity				
	Mean Score	Median	Diff	Diff Range	Confidence Level
MoDOT AFAD	8.94	10	Baseline		
AFAD with Alternative Sign	8.56	9	0	(-1, 0)	79.80%
AFAD without CMS	6.13	6	-3	(-4, -2)	> 99.9%*
Flagger	6.41	6	-2	(-4, 0)	> 99.9%*
	Visibility				
	Mean Score	Median	Diff	Diff Range	Confidence Level
MoDOT AFAD	9.47	10	Baseline		
AFAD with Alternative Sign	8.44	9	-1	(-2, 0)	99.9%*
AFAD without CMS	6.41	6	-3	(-4, -2)	> 99.9%*
Flagger	4.19	4	-5	(-7, -5)	> 99.9%*
	Safety				
	Mean Score	Median	Diff	Diff Range	Confidence Level
MoDOT AFAD	9.19	10	Baseline		
AFAD with Alternative Sign	8.5	8.5	-1	(-1, 0)	98.6%**
AFAD without CMS	6.28	6.5	-3	(-4, -2)	> 99.9%*
Flagger	4.06	4	-5	(-6, -4)	> 99.9%*
	Efficiency				
	Mean Score	Median	Diff	Diff Range	Confidence Level
MoDOT AFAD	8.74	9	Baseline		
AFAD with Alternative Sign	8.48	9	0	(-1, 0)	86.40%
AFAD without CMS	6.58	6	-2	(-3, -1)	> 99.9%*
Flagger	5.29	5	-3	(-5, -2)	> 99.9%*

* indicates significance at 99% confidence level
** indicates significance at 95% confidence level

In response to Question 7 regarding the necessity of the CMS, 78 percent of participants agreed or strongly agreed that the CMS was necessary for the AFAD as shown in Table 4.1.4.6. Half of the participants strongly agreed that the CMS was necessary while only 10 percent of participants disagreed or strongly disagreed that the CMS was necessary. The results from this question support the use of the CMS in conjunction with the AFAD.

Table 4.1.4.6 Survey Results for Necessity of CMS

CMS Necessary?	Count	Percentage	
Strongly Agree	16	50.00%	78.13%
Agree	9	28.13%	
Neutral	4	12.50%	12.50%
Disagree	1	3.13%	9.38%
Strongly Disagree	2	6.25%	

Questions 8 and 9 in Part 3 of the survey evaluated the fidelity of the simulator, and the results are shown in Table 4.1.4.7. Among the participants, 75 percent of them agreed or strongly agreed that they felt like they were driving on a real highway, and 81 percent of them agreed or strongly agreed that they felt like they could drive freely. Approximately 6 percent of participants disagreed or strongly disagreed with the highway fidelity of the simulator, and less than 10 percent of participants disagreed with the feeling of driving freely.

Table 4.1.4.7 Survey Results for Simulator Fidelity

	Fidelity of Highway			Drive Freely		
	Count	Percentage		Count	Percentage	
Strongly Agree	12	37.50%	75.00%	11	34.38%	81.25%
Agree	12	37.50%		15	46.88%	
Neutral	6	18.75%	18.75%	3	9.38%	9.38%
Disagree	1	3.13%	6.25%	3	9.38%	9.38%
Strongly Disagree	1	3.13%		0	0.00%	

Participants provided written comments on the different designs and the overall simulation experience. Some thought the human flagger was unsafe (five participants) and the flagger was less visible than AFADs (eight participants). Two participants said that they would be more careful when seeing a human flagger, and three thought that it was easy to understand a flagger. Participants commented that MoDOT AFAD was highly visible (six participants), safe (one participant), clear (three participants), and love it (three participants). Although one participant commented that the MoDOT AFAD was easy to understand, two thought that the MoDOT AFAD would be confused with a STOP sign. Some said the AFAD with alternative sign was safe (one participant), and provided clearer instructions (one participant), but the letters on the CMS was too small (four participants). Two participants were confused by it. Seven participants commented that the AFAD without CMS was confusing, of which three thought that the blank CMS board was more confusing. Three of them thought its visibility was not high enough. Two participants thought the simulator was great.

Other comments and suggestions included:

- Add more road signs before the human flagger.
- Put more flash lights on AFADs to distinguish it from cars.
- AFADs consumed more energy, solar powered may be more efficient.
- The white sign showing “WAIT ON STOP” was hard to read from far distance. It was confusing and should be removed.
- Add night time scenarios for the test.

SSQ Results

The simulator experiment was followed by a 16-question SSQ, asking if participants felt sick during or after the simulator experiment. The results are shown in Table 4.1.4.8, and the percentages are presented in Appendix H. Most participants felt no or slight discomfort, and only three out of 32 participants had moderate or worst symptoms.

Among the three participants with moderate or severe symptoms, one participant felt moderate general discomfort, dizziness with eyes open, stomach awareness, and severe nausea. The second one felt moderate eye strain, difficulty focusing, and fullness of the head. The last one felt moderate nausea and stomach awareness.

Table 4.1.4.8 SSQ Results

	General discomfort	Fatigue	Headache	Eye strain
None	25	31	24	23
Slight	6	1	8	8
Moderate	1	0	0	1
Severe	0	0	0	0
	Difficulty focusing	Salivation increasing	Sweating	Nausea
None	27	31	30	26
Slight	4	1	2	4
Moderate	1	0	0	1
Severe	0	0	0	1
	Difficulty concentrating	Fullness of the Head	Blurred vision	Dizziness with eyes open
None	31	27	28	27
Slight	1	3	3	4
Moderate	0	1	0	1
Severe	0	0	0	0
	Dizziness with eye closed	Vertigo	Stomach awareness	Burping
None	30	31	27	30
Slight	2	1	3	2
Moderate	0	0	2	0
Severe	0	0	0	0

4.1.5 Conclusions

The field test, simulator study, and multiple survey results were consistent in showing that the MoDOT AFAD performed better than human flaggers using multiple MOEs. The results indicated that the AFAD may enhance safety over the human flagger based on a reduced vehicle approach speed, farther full stop location, and lower noncompliance rate. The public had a favorable impression of the AFAD and generally preferred it over the human flagger. Among all AFADs, the MoDOT AFAD had the most outstanding performance and was preferred the most.

One possible explanation for the results is that the TMA truck and CMS increased the visibility of the AFAD, which helped to reduce approach speeds and increase stopping distances. The combination of the STOP/SLOW paddle with the red/yellow lights (MUTCD option) also helped drivers better understand the device. Besides aforementioned results, another benefit of the AFAD is that work zone operator sits inside the truck and thus protected by the truck and TMA, while human flaggers are exposed directly to traffic. These results are highly encouraging for any jurisdictions who are interested in pursuing the use of AFADs to improve work zone and worker safety.

These promising results should be interpreted with some issues in mind. Despite the similar trends shown in field and simulator studies, the absolute magnitudes of MOEs differed. This illustrates the fact the simulator studies are better at establishing relative validity than absolute validity (Kaptein et al. 1996). The results were obtained from work zones on a rural highway; results may be different in urban area. The impacts of other factors, like traffic volume, lane closure length, and speed limit, were not examined in this study. All MOEs were obtained from drivers in Missouri, and all AFADs device were new to them. Therefore, the results of AFADs on driver behaviors may vary in a different state, and the novelty effect of AFAD designs should be examined in a study of longer duration.

Besides safety feature, the cost of an AFAD is another factor to be considered for implementation. According to a staff summary by MoDOT (MoDOT 2008), the cost of a pair of AFAD device ranged from $14,999 to $30,640, with additional cost of trailers. During an in-person meeting, a MoDOT staff mentioned that the cost of MoDOT AFAD was around $36,000. A new MoDOT typical changeable message sign (CMS) costed around $12,000 (MoDOT 2017). Comparing the cost of an AFAD to a human flagger, which costed labor, insurance, and personal protection equipment, an AFAD seemed to be expensive. But life is priceless. Based on the project conclusions and the tradeoff between price and safety, MoDOT decided to deploy MoDOT AFAD to work zones statewide in Fall 2017.

The MoDOT AFAD is suitable for stationary work zones on two-way two-lane rural highways with relatively low speed (lower than 65 mph), as vehicles have to come to a full stop before reaching the AFAD. However, under a high-speed driving environment, especially when drivers are aggressive, AFAD may not be as effective as possibility of crash may increase. The MoDOT AFAD protects operator well as operator sit in the truck and controls traffic. It has good safety feature as the AFAD is mounted on a truck-mounted attenuator (TMA), even if it gets hit, the attenuator acts like a buffer zone and reduces the impact of crash. It is transferrable to other states, as components of the MoDOT AFAD are easily understandable and is not very likely to cause misunderstanding due to discrepancy of understanding between public in different states. Also, the whole AFAD unit is mounted on a truck, it has high mobility and could be transported easily. The mounting is removeable, which provides extra flexibility.

4.2 GREEN LIGHTS ON TMAS

Truck-mounted attenuators (TMAs) are designed to improve mobile work zone safety by shadowing the working truck, enhancing visibility of a work zone, and catching drivers’ attention early to slow them down when driving through a mobile work zone. Despite the use of the TMAs and other precautions such as shadow vehicles, arrow boards, and signs to warn drivers that they are approaching a mobile work zone, some drivers do not respond to warnings and collide with the TMA. The goal of this research project was to help improve safety in mobile work zones. The objective of this project was to evaluate the effectiveness of green lights on TMAs and determine the best TMA light bar configuration. The scope of the project included two phases: a simulator test with four configurations and a field test with two configurations. The simulator testing phase examined amber/white ("MoDOT typical"), green only ("MoDOT preferred"), green/amber ("MoDOT alternative"), and green/white (Design alternative) configurations. In the field test, only amber/white and green only configurations were evaluated. The simulator testing phase was composed of three elements: the regular simulator scenario followed by a post-simulator survey, a disability glare test utilizing an eye tracker, and a visibility test. Detailed quantitative measures were used for the first time in this study to evaluate green lights on TMAs in the United States.

An example of the TMA configuration is shown in Figure 4.2.0.1. The TMA was equipped with a flashing light bar on top of the arrow board and lights on top of the checkerboard to enhance visibility. This project mainly focused on the evaluation of the light colors on the top light bar.

Figure 4.2.0.1 Example of green only TMA

4.2.1 Simulator Study

Phase One of the project consisted of the simulator study. The University of Missouri's driving simulator, ZouSim, was used to examine four different TMA lighting configurations in a cost-effective way. The simulator provided safe experiment conditions for the testing of alternatives in a highly controlled environment that limited extraneous causal factors.

Simulator Setup

The study simulated work zones on a divided, four-lane freeway in Missouri with a speed limit of 60 mph. The entire highway segment was designed without vertical or horizontal curves, in order to eliminate the influence of terrain. The pavement was created based on AASHTO Green Book standards (AASHTO 2013). Surfaces were textured and/or painted with the appropriate striping and markings that conform to the MUTCD (FHWA 2009). The work zone layout is shown in Figure 4.2.1.1. In the right lane closed work zone, all work zone vehicles were moving at 10 mph. The work zone configuration followed the requirements of MoDOT Engineering Policy Guide (MoDOT 2018).

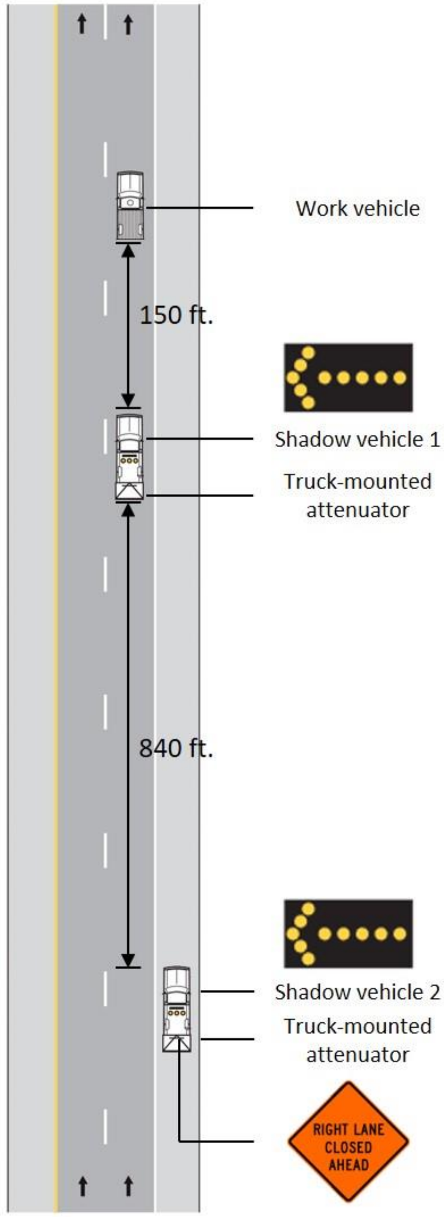


Figure 4.2.1.1 Mobile work zone layout in simulator tests

Four TMA light bar configurations were tested in the simulator tests: (1) amber/white (MoDOT typical), (2) green only (MoDOT preferred), (3) green/amber (MoDOT alternative), and (4) green/white (design alternative). There was no other traffic, except the work zone vehicles in the direction of travel, so the participant was free to accelerate/decelerate or change lanes. Therefore, all the measurements derived from drivers were responding to TMAs only.

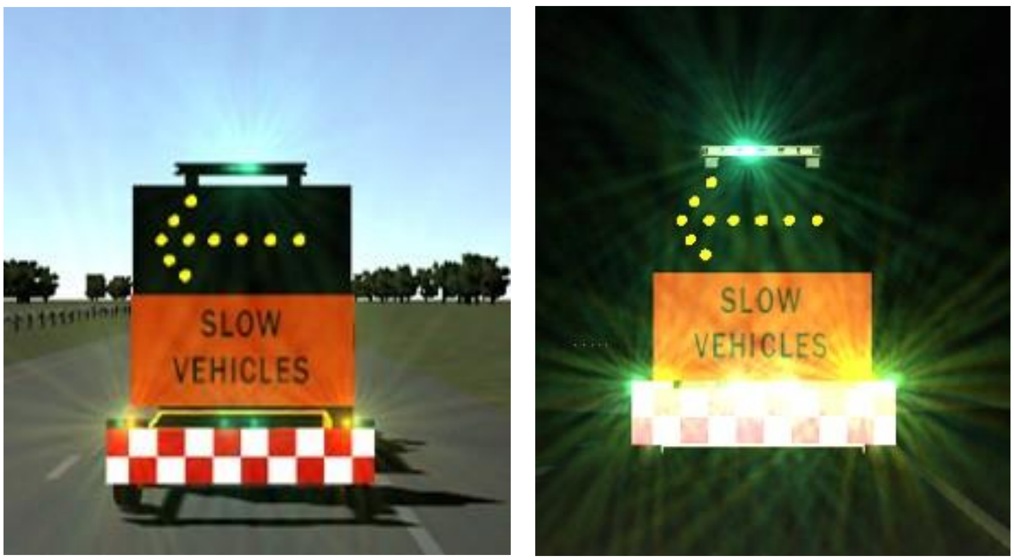
Each configuration was tested for both daytime and nighttime scenarios. The nighttime scenarios used the same settings as the daytime scenarios, except that lights were dimmed by half. An eye tracker was utilized to capture the movement of participants' pupils for nighttime scenarios, as high intensity of lights could cause disability glare issues. Disability glare refers to a situation where the light bar could be so dominant that drivers might be unable to see the flashing arrow and other items near the light bar. A water bottle rolled next to the rear advance TMA, and the eye tracker indicated if participants saw the water bottle, so as to determine disability glare of each configuration. The water bottle was chosen as the least obtrusive and yet measurable option after considering several other options.

The amber and white configuration was the baseline for the experiment. The amber and white TMA was equipped with amber and white light bars on top of arrow board, an arrow board with the arrow flashing, and amber and white light bulbs on the checkerboard, as shown in Figure 4.2.1.2.

(a) (b)

Figure 4.2.1.2 amber/white scenarios (a) daytime (b) nighttime

The green only TMA was preferred by MoDOT based on field testing that was performed at Lee's Summit airport by MoDOT. The feedback from the pilot project crew was positive. The crew thought that the green only TMA worked well to attract driver attention, and the visibility was high. TMA operators felt safer with green lights. However, concerns about confusion with traffic signal and arrow board were expressed. The green only TMA was equipped with green light bars on top of arrow board, an arrow board with arrow flashing, and green and amber light bulbs on the checkerboard, as shown in Figure 4.2.1.3.



(a) **(b)**

Figure 4.2.1.3 Green only scenarios (a) daytime (b) nighttime

The green and amber was an alternative configuration tested by MoDOT at Lee's Summit airport. The simulated configuration was the same as the one tested by MoDOT, with green/amber light bars on top of the arrow board and green/amber light bulbs on the checkerboard, as shown in Figure 4.2.1.4.

(a) (b)

Figure 4.2.1.4 Green/amber scenarios (a) daytime (b) nighttime

The research team designed an alternative green/white TMA, with green/white light bars on top of arrow board and amber/green light bulbs on the checkerboard, as shown in Figure 4.2.1.5. The reasons why research team chose green and white were: 1) amber, green and white colors were tested in the other three configurations, and green was the major color to be tested, combining white with green was natural; 2) optically, white was the color that contained all other colors, and it was commonly used in daily life. If white was effective, then the adjustment to be made to TMA light bars would be easy.

(a) (b)

Figure 4.2.1.5 Green/white scenarios (a) daytime (b) nighttime

In any simulator study, sequence bias or order effect is possible (Perreault 1975). In order to limit this bias, the test order of scenarios was randomized. Daytime scenarios and nighttime scenarios were randomized separately and combined together randomly. Four configuration scenarios generated 24 orders. Daytime and nighttime scenarios were paired randomly, therefore generating 576 combinations. Each participant was randomly assigned to one of the 576 test orders to minimize sequence bias. Each of the four TMA configurations was tested once in daytime and once in nighttime.

Simulator Trials and MOEs

The campus Institutional Review Board reviewed, evaluated, and approved the study protocols and measurement tools. A standard trial hosting script was used for each trial. The host first introduced the simulator, described the purpose of the experiment, and obtained participant consent for data collection. Then the participant sat in the sedan cab, and the host calibrated the eye tracker to capture participant eye gaze. After calibration, the participant practiced driving the simulator in a free-driving warm-up scenario to become familiar with the simulator. Once the participant felt comfortable driving the simulator, the host initiated the actual work zone scenarios, beginning with the daytime scenarios. Once the daytime scenarios were completed, the nighttime scenarios were initiated. The participant was asked to drive along the highway and stay in the driving lane (right lane) as much as possible since the lane change distance and speed were used as performance measures. Otherwise, a driver might stay in the passing lane for the entire simulator run.

There were three different parts of the simulator tests. The first part was regular simulator trials in which drivers traveled through the work zone under the different scenarios and information regarding speeds and distances was collected. The second part of the simulator testing was the disability glare test in which the eye tracker was utilized to capture participants' eyes and examine the disability glare of each light configuration. This test was conducted as part of the regular simulator trials. The third part of the simulator testing was the visibility test in which the participant would push a button when he/she first recognized the work zone and arrow direction, separately. The visibility test was conducted apart from the regular simulator trials to avoid distracting participants when they were driving.

The simulator trials, including eye tracking, were all recorded. After simulator experiments were completed, the research team reviewed the videos and extracted data. Nine measures of effectiveness (MOEs) were defined for data reduction.

Regular trials

In this part of the simulator testing, six MOEs were captured. These MOEs were measured for both daytime and nighttime scenarios.

- MOE 1: - First blinker distance (ft.). This MOE is the distance to the shoulder TMA when the participant flashed the blinker for the first time. This distance could help to indicate where a driver recognized the work zone and/or recognized the arrow direction. An example of MOE 1 is shown in Figure 4.2.1.6. The figure shows the green blinker indicator that flashes on the vehicle's heads up display and the distance superimposed on the video. For illustration purposes, Figure 4.2.1.6 and subsequent figures show a composite view of the front windshield view and right-side mirror view. Although the side view appears to be distorted in these figures, the rear view and side mirrors are situated properly from the perspective of sitting inside the vehicle cab.

Figure 4.2.1.6 MOE 1 example: first blinker distance (ft.)

- MOE 2: First blinker speed (mph). This MOE is the speed when the participant flashed the blinker for the first time. The speed is available on the heads up display as shown in Figure 4.2.1.7. Speed is a common effectiveness measure that is related to safety.

Figure 4.2.1.7 MOE 2 example: first blinker speed (mph)

- MOE 3: Merge distance (ft.). This MOE is the distance to the shoulder TMA when the participant started to merge to the passing lane. This distance indicates how far upstream the participant reacted to the work zone and/or arrow board. An example of MOE 3 is shown in Figure 4.2.1.8.

Figure 4.2.1.8 MOE 3 example: merge distance (ft.)

- MOE 4: Merge speed (mph). This MOE is the speed of vehicle when it started to merge to the passing lane. An example of MOE 4 on the screen is shown in Figure 4.2.1.9.

Figure 4.2.1.9 MOE 4 example: merge speed (mph)

- MOE 5: Speed when passing the shoulder TMA (mph). This MOE is the speed of the vehicle when it passed the rear end of the shoulder TMA. A lower speed is desirable for safety. An example of MOE 5 is shown in Figure 4.2.1.10.

Figure 4.2.1.10 MOE 5 example: speed when passing shoulder TMA (mph)

- MOE 6: Speed when passing the rear advance TMA (mph). This MOE is the speed of vehicle when it passed the rear end of rear advance TMA. A lower speed is more desired for safety consideration. An example of MOE 6 is shown in Figure 4.2.1.11.

Figure 4.2.1.11 MOE 6 example: speed when passing rear advance TMA (mph)

<u>Disability glare test</u>

The light bars above the arrow board could potentially impact arrow visibility due to light intensity and may cause disability glare, especially during nighttime. The simulator study was used to assess possible disability glare and discomfort effects. A water bottle was rolling right next to the rear advanced TMA. The eye tracker was utilized to track a participant's eyes and to determine if the participant saw the water bottle. The eye tracker was shown as a bubble on the monitoring screen of the host, but not shown on the driver's display monitor to avoid distraction. The following MOE was used to assess disability glare:

- MOE 7: Water bottle recognition is a binary measurement, indicating if the participant saw the water bottle or not. This MOE is for nighttime only. Figure 4.2.1.12 shows an example of the eye tracker indicating that water bottle was recognized.

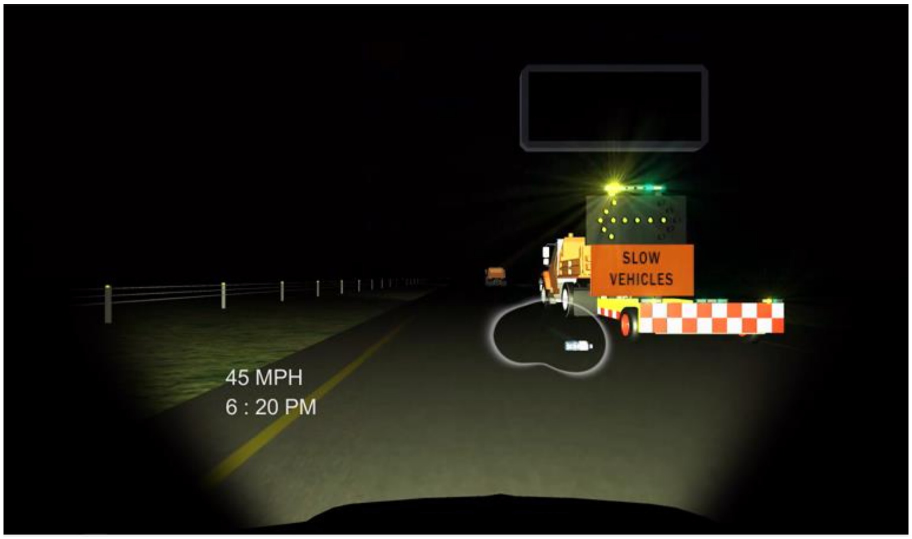


Figure 4.2.1.12 MOE 7 example: water bottle recognition

<u>Visibility test</u>

In order to measure the visibility of mobile work zones with different light configurations without distracting the participants, a separate visibility test was conducted. Although visibility of the work zone and arrow board could be implied from first blinker and merge distance, a separate test was performed to capture the recognition distance directly. During this test, participants were asked to press two buttons. The first button was pressed when the participants realized that there was a moving work zone ahead, based on the visibility of light bar. The second button was pressed when they could tell what direction the arrow is pointing, to test the visibility of the arrow aboard. The simulator logged the distance measurements automatically when participants pressed the buttons. For the visibility testing, twenty participants were invited to drive through the four light bar configurations twice (daytime/nighttime). Two MOEs were generated from this test as described below.

- MOE 8: Work zone recognition distance (ft.). This MOE is the distance from the rear end of the shoulder TMA when the participant first recognized the work zone. A farther distance indicated that the work zone was more visible and could be safer because of longer reaction times.
- MOE 9: Arrow direction recognition distance (ft.). This MOE is the distance from the rear end of the shoulder TMA when the participant first recognized the arrow direction. This could be an indication of disability glare as well. A farther distance is desirable.

Both MOE 8 and MOE 9 were automatically recorded onto a data log when the buttons were pushed.

Simulator Results

Videos of regular trials and the disability glare test were recorded, and MOEs 1 through 7 were extracted. Video was useful to visually confirm data accuracy and identify potential data issues. The data for MOE 8 and MOE 9 of the visibility test were extracted from automated data files because they were very straight forward and did not require visual confirmation and interpretation.

To compare MOEs among the TMA light bar configurations, the differences between them were calculated. Mean and standard deviation (Stdev.) of MOEs were reported. Statistical analysis was performed to calculate significance, confidence level, and effect size. A confidence level higher than 90 percent was regarded as significant in this study. Effect size was presented as Cohen's d value, and difference is defined to be small (Cohen's $d < 0.5$), medium (Cohen's d between 0.5 and 0.8), and large (Cohen's $d > 0.8$) (Cohen 1977).

Thirty participants participated in the first part of the simulator tests and completed the trials. However, distance data for one of the participants was lost due to a data issue. MOE 1 and MOE 2 data were not obtained for some participants because they did not flash their blinker before they merged.

The eye tracker was utilized for 18 participants because it became available during the middle of the simulator testing. Due to contact lens issues, some eye tracking calibrations were not successful and eye tracking data was not collected.

The visibility test was conducted separately after the simulator trials because it would be distracting if participants were asked to push a button when they first recognized the work zone and arrow direction while they were driving in a normal manner. The objective for the visibility test was to distinguish the configuration with the earliest recognition of the work zone and arrow board. Twenty participants participated in the visibility test. Some of them were participants from the previous simulator tests, and the others were newly recruited.

Daytime results

MOE 1 measured the distance of a vehicle from the rear end of the shoulder TMA when the participant flashed the blinker for the first time. As shown in Table 4.2.1.1, participants reacted to amber/white TMA the quickest among all four configurations and flashed the blinker at an average distance of 1039.1 ft. Both the green only and green/white TMA had significantly closer first blinker distances than the amber/white TMA, but Cohen's d showed that the effect sizes were small.

Table 4.2.1.1 Daytime MOE 1: First blinker distance (ft.)

	MOE 1: First blinker distance (ft.)						
	Count	**Mean**	**Stdev.**	**Diff.**	**Confidence level**	**Cohen's d**	**Effect size**
Amber/white	24	1039.1	589.2	Baseline			
Green only	24	813.8	377.7	-225.3	97.6%**	0.46	Small
Green/amber	24	973.1	508.6	-66.0	73.1%	n/a	n/a
Green/white	25	788.4	444.4	-250.7	99.8%*	0.48	Small

* indicates significance at 99%.
** indicates significance at 95%.

MOE 2 captured the speed of a vehicle when the participant flashed the first blinker, and the results are shown in Table 4.2.1.2. The amber and white had the highest first blinker speed, the results being statistically significant. However, the Cohen's d indicated that the differences were small. The reason why the speed for amber/white TMA was higher could be that the speed was captured when vehicle was farther away from the TMA compared to other light configurations.

Table 4.2.1.2 Daytime MOE 2: First blinker speed (mph)

	MOE 2: First blinker speed (mph)						
	Count	**Mean**	**Stdev.**	**Diff.**	**Confidence level**	**Cohen's d**	**Effect size**
Amber/white	25	57.8	7.6	Baseline			
Green only	25	55.7	7.7	-2.0	91.5%***	0.27	Small
Green/amber	25	54.8	8.6	-3.0	96.7%**	0.36	Small
Green/white	26	54.8	9.1	-3.0	92.3%***	0.35	Small

** indicates significance at 95%.
*** indicates significance at 90%.

Since the rear advance TMA was on the driving lane, it is important to capture the merge distance of vehicles as a surrogate measure of safety. The merge distance was recorded as MOE 3, and the results for this MOE are shown in Table 4.2.1.3. The results indicate that drivers merged earlier when they encountered the amber/white TMA, followed by green/amber TMA. The merge distance for the green only TMA was significantly closer than the amber/white TMA, and the effect size was medium. It could be implied that the visibility of the green only TMA was lower and drivers reacted later.

Table 4.2.1.3 Daytime MOE 3: Merge distance (ft.)

	MOE 3: Merge distance (ft.)						
	Count	**Mean**	**Stdev.**	**Diff.**	**Confidence level**	**Cohen's d**	**Effect size**
Amber/white	29	971.0	599.5	Baseline			
Green only	29	649.0	336.6	-322.0	99.7%*	0.66	Medium
Green/amber	28	814.5	468.2	-156.5	83.2%	n/a	n/a
Green/white	29	750.1	643.4	-220.9	98.0%**	0.36	Small

* indicates significance at 99%.
** indicates significance at 95%.

Merge speed was recorded as well as MOE 4. Table 4.2.1.4 shows that the merge speed for the amber/white TMA was higher than the other light configurations. This may be due to the farther merge distance, while drivers had not slowed down yet.

Table 4.2.1.4 Daytime MOE 4: Merge speed (mph)

	MOE 4: Merge speed (mph)						
	Count	**Mean**	**Stdev.**	**Diff.**	**Confidence level**	**Cohen's d**	**Effect size**
Amber/white	30	57.4	6.5	Baseline			
Green only	30	53.7	11.4	-3.7	95.1%**	0.40	Small
Green/amber	30	53.5	10.4	-3.9	96.2%**	0.45	Small
Green/white	30	53.0	8.5	-4.4	98.3%**	0.57	Medium

** indicates significance at 95%.

MOE 5 measured the speed of the vehicle when it passed the shoulder TMA. The results for this MOE are shown in Table 4.2.1.5. The results show that the green only TMA had the highest passing speed, and the green/white TMA had the lowest passing speed which is desirable. However, the differences among all four configurations were not statistically significant.

Table 4.2.1.5 Daytime MOE 5: Speed when passing shoulder TMA (mph)

	MOE 5: Speed when passing shoulder TMA (mph)						
	Count	**Mean**	**Stdev.**	**Diff.**	**Confidence level**	**Cohen's d**	**Effect size**
Amber/white	30	52.3	9.4	Baseline			
Green only	30	53.2	9.3	0.9	81.5%	n/a	n/a
Green/amber	30	52.1	9.4	-0.2	55.0%	n/a	n/a
Green/white	30	50.5	8.9	-1.8	87.7%	n/a	n/a

MOE 6 is similar to MOE 5, and it measured the vehicle speed when it passed the rear advance TMA. The results for this MOE are shown in Table 4.2.1.6. The green only TMA had the highest passing speed, and green/amber TMA had the lowest. However, the speeds for all four configurations were close to each other, and they were not statistically significantly different from each other.

Table 4.2.1.6 Daytime MOE 6: Speed when passing rear advance TMA (mph)

	MOE 6: Speed when passing rear advance TMA (mph)						
	Count	**Mean**	**Stdev.**	**Diff.**	**Confidence level**	**Cohen's d**	**Effect size**
Amber/white	30	51.0	8.3	Baseline			
Green only	30	52.1	9.6	1.2	89.1%	n/a	n/a
Green/amber	30	50.6	9.6	-0.4	63.0%	n/a	n/a
Green/white	30	50.8	8.5	-0.2	57.3%	n/a	n/a

The second part of the simulator tests was the disability glare test. The eye tracker was utilized, and results were presented as binary data to indicate whether or not the participant recognized the water bottle. MOE 7, water bottle recognition, was the only measurement in the disability glare test. This MOE was assessed for nighttime only because the disability glare issue was not considered as a critical factor during daytime. Therefore, the daytime results did not contain MOE 7. The nighttime results for this MOE are presented following the daytime results.

Table 4.2.1.7 shows the results of work zone recognition distance (MOE 8) in the daytime. Since the sample size of the visibility testing was smaller than the regular simulator trials, the statistical tests conducted were different. A Mann-Whitney test was conducted to assess the statistical significance of the data. The Mann-Whitney test is used to assess data that is not normally distributed by calculating average score differences and determining if difference between the data sets is significant (De Winter and Dodou 2010). The results show that the amber/white configuration, as the baseline in this study, had the farthest distance with a mean of 1554 ft. and median of 1604 ft. Meanwhile, the visibility of the green only TMA was the lowest (1166 ft.) among all configurations. However, based on the results of the Mann-Whitney test, there was no significant difference between these four configurations.

Table 4.2.1.7 Daytime MOE 8: Work zone recognition distance (ft.)

	MOE 8: Work zone recognition distance (ft.)					
	Count	**Mean**	**Median**	**Stdev.**	**Diff.**	**Confidence Level**
Amber/white	20	1554.2	1604.0	746.2	Baseline	
Green only	20	1166.2	1154.5	469.6	388.0	90.9%
Green/amber	20	1292.8	1311.5	546.0	261.4	69.0%
Green/white	20	1382.6	1231.0	722.6	171.6	59.1%

Participants were also asked to state when they could recognize the arrow direction, and the results of this test for daytime are shown in Table 4.2.1.8 as MOE 9. Even though the green only TMA scored the lowest for work zone recognition distance, it had the farthest distance for recognition of the arrow direction with a mean of 663 ft. and median of 567 ft. In contrast, the configuration of amber and white had the nearest distance (569.70 ft.), but all the differences were not statistically significant among four configurations.

Table 4.2.1.8 Daytime MOE 9: Arrow direction recognition distance (ft.)

	MOE 9: Arrow direction recognition distance (ft.)					
	Count	**Mean**	**Median**	**Stdev.**	**Diff.**	**Confidence Level**
Amber/white	20	569.7	522.5	195.2	Baseline	
Green only	20	663.0	567.0	224.8	93.3	81.9%
Green/amber	20	644.6	547.0	215.8	74.9	73.3%
Green/white	20	578.6	553.0	137.6	9.6	29.5%

Nighttime results

MOE 1 measured the distance of the vehicle from the rear end of the shoulder TMA when the participant flashed the blinker for the first time, and the results were different from daytime. As shown in Table 4.2.1.9, participants reacted to the amber/white TMA at 1577.6 ft. which was farther than the same measurement for daytime. It was also the earliest reaction among all four configurations, followed by the green/amber TMA. Both the green only and green/white TMA had significantly closer first blinker distances than the amber/white TMA, and the Cohen's d showed that the effect sizes were large. It can be implied that the amber/white had significantly higher visibility than the green only TMA and green/white TMA during nighttime. Compared to daytime, participants reacted to the TMA earlier at night.

Table 4.2.1.9 Nighttime MOE 1: First blinker distance (ft.)

	MOE 1: First blinker distance (ft.)						
	Count	**Mean**	**Stdev.**	**Diff.**	**Confidence level**	**Cohen's d**	**Effect size**
Amber/white	24	1577.6	748.1	Baseline			
Green only	23	869.1	389.7	-708.5	99.9%*	1.19	Large
Green/amber	23	1483.0	1218.3	-94.6	66.0%	n/a	n/a
Green/white	23	887.3	462.8	-690.3	99.9%*	1.11	Large

* indicates significance at 99%.

MOE 2 captured the speed of the vehicle when the participant flashed the first blinker, and the results are shown in Table 4.2.1.10. The overall speed at first blinker at nighttime was higher than daytime. This higher speed may have occurred because the vehicle was located farther away from the work zone. However, although the first blinker distance of the amber/white TMA was the farthest, the speed was not the highest. This result may indicate that amber/white caught drivers' attention and alerted them. The green only TMA had a slightly lower first blinker speed, but the difference was not significant.

Table 4.2.1.10 Nighttime MOE 2: First blinker speed (mph)

	MOE 2: First blinker speed (mph)						
	Count	**Mean**	**Stdev.**	**Diff.**	**Confidence level**	**Cohen's d**	**Effect size**
Amber/white	25	58.3	5.9	Baseline			
Green only	24	57.4	5.6	-0.9	82.7%	n/a	n/a
Green/amber	24	59.5	5.5	1.2	77.8%	n/a	n/a
Green/white	24	59.6	4.2	1.3	85.0%	n/a	n/a

The merge distances at nighttime are shown in Table 4.2.1.11. The results indicate that drivers merged earlier when they encountered the amber/white TMA and green/amber TMA. Both the green only TMA and green/white TMA had significantly closer merge distances, and the effect sizes were large and medium correspondingly. Similar to the first blinker distance, drivers reacted to TMAs earlier during nighttime and merged earlier.

Table 4.2.1.11 Nighttime MOE 3: Merge distance (ft.)

	MOE 3: Merge distance (ft.)						
	Count	**Mean**	**Stdev.**	**Diff.**	**Confidence level**	**Cohen's d**	**Effect size**
Amber/white	29	1309.4	762.3	Baseline			
Green only	29	706.5	398.1	-602.9	99.9%*	0.99	Large
Green/amber	29	1288.7	1134.5	-20.8	56.1%	n/a	n/a
Green/white	29	851.9	603.6	-457.5	99.8%*	0.67	Medium

* indicates significance at 99%.

Merge speed at nighttime was slightly higher than at daytime, a result which may be due to farther merge distance at nighttime. As shown in Table 4.2.1.12, the green only TMA and amber/white TMA had slower merge speeds than the other two configurations. However, the differences were not significant. With regard to the merge distance and speed, the green/white TMA did not perform as well as the other three configurations, and this was consistent with the results from first blinker distance and speed.

Table 4.2.1.12 Nighttime MOE 4: Merge speed (mph)

	MOE 4: Merge speed (mph)						
	Count	**Mean**	**Stdev.**	**Diff.**	**Confidence level**	**Cohen's d**	**Effect size**
Amber/white	30	57.3	6.6	Baseline			
Green only	30	56.0	6.2	-1.2	82.3%	n/a	n/a
Green/amber	30	57.9	5.7	0.6	66.0%	n/a	n/a
Green/white	30	58.2	4.5	0.9	77.2%	n/a	n/a

MOE 5 was the measurement of speed when the vehicle drove past the shoulder TMA. The results of MOE 5 are shown in Table 4.2.1.13. It shows that the green/amber TMA may have caught drivers' attention the most, as it had the lowest speed, and the speed was statistically significantly slower than the speed under the baseline configuration. The speed for the amber/white TMA was the highest. This result may be due to the discomfort brought by the brightness of the amber and white light bars.

Table 4.2.1.13 Nighttime MOE 5: Speed when passing shoulder TMA (mph)

	MOE 5: Speed when passing shoulder TMA (mph)						
	Count	**Mean**	**Stdev.**	**Diff.**	**Confidence level**	**Cohen's d**	**Effect size**
Amber/white	30	51.8	9.5	Baseline			
Green only	30	50.9	10.4	-0.9	74.9%	n/a	n/a
Green/amber	30	48.6	10.1	-3.2	98.2%**	0.33	Small
Green/white	30	51.1	8.9	-0.7	70.9%	n/a	n/a

** indicates significance at 95%.

The results for the speed of vehicles when they passed the rear advance TMA are shown in Table 4.2.1.14 as MOE 6. The passing speeds for the rear advance TMA were slower than the passing speeds for the shoulder TMA. This result may have occurred because the rear advance TMA was on the lane and thus closer to vehicles, while the shoulder TMA was on shoulder. The green and amber TMA had the slowest passing speed. However, the differences between the four configurations were not significant.

Table 4.2.1.14 Nighttime MOE 6: Speed when passing rear advance TMA (mph)

	MOE 6: Speed when passing rear advance TMA (mph)						
	Count	**Mean**	**Stdev.**	**Diff.**	**Confidence level**	**Cohen's d**	**Effect size**
Amber/white	30	48.6	11.2	Baseline			
Green only	30	49.3	10.4	0.8	73.0%	n/a	n/a
Green/amber	30	47.7	9.6	-0.9	72.9%	n/a	n/a
Green/white	30	48.2	9.4	-0.4	61.2%	n/a	n/a

The disability glare test was used to assess if the light bars were too bright and caused glare issues when drivers drove past the TMA. By using a water bottle next to rear advance TMA and utilizing the eye tracker to determine if participants saw the water bottle, the presence of glare could be implied. As shown in Table 4.2.1.15, the results indicate that the amber/white TMA had the lowest rate of water bottle recognition, while the green/white TMA had the highest rate of water bottle recognition. Thus, the amber/white configuration exhibited more problems with glare than the other configurations.

Table 4.2.1.15 MOE 7: Water bottle recognition (binary)

	MOE 7: Water bottle recognition		
	Count	**Recognized**	**Percentage**
Amber/white	14	10	71.4%
Green only	12	11	91.7%
Green/amber	13	11	84.6%
Green/white	13	12	92.3%

The visibility test was also performed in nighttime. Table 4.2.1.16 for MOE 8 shows the results for participants' recognition of the moving work zone. The amber/white TMA had the farthest recognition distance with a mean of 1388.7 ft. and median of 1390 ft. The green only TMA had the closest recognition mean distance (976.5 ft.), and the difference between it and baseline was statistically significant.

Table 4.2.1.16 Nighttime MOE 8: Work zone recognition distance (ft.)

	MOE 8: Work zone recognition distance (ft.)					
	Count	**Mean**	**Median**	**Stdev.**	**Diff.**	**Confidence Level**
Amber/white	20	1388.7	1390.0	419.0	Baseline	
Green only	20	976.5	935.5	319.9	412.2	99.6% *
Green/amber	20	1162.5	1155.5	473.3	226.2	89.8%
Green/white	20	1296.0	1154.0	667.8	92.7	67.7%

* indicates significance at 99%.

The results for arrow direction recognition at nighttime are shown in Table 4.2.1.17. The results show that the green only TMA had the farthest arrow direction recognition distance with a mean of 555.1 ft. and median of 552.0 ft., a result which was consistent with the daytime findings. However, there was not a statistically significant difference among the four configurations.

Table 4.2.1.17 Nighttime MOE 9: Arrow direction recognition distance (ft.)

	MOE 9: Arrow direction recognition distance (ft.)					
	Count	**Mean**	**Median**	**Stdev.**	**Diff.**	**Confidence Level**
Amber/white	20	530.9	500.0	156.4	Baseline	
Green only	20	555.1	552.0	161.5	24.2	78.0%
Green/amber	20	546.9	517.0	110.6	16.0	54.3%
Green/white	20	519.5	479.0	142.5	11.4	2.3%

The results of MOE 8 and MOE 9 imply that the recognition of work zone and recognition of arrow direction are negatively correlated. In other words, a more intense light that leads to the earlier recognition of the mobile work zone also means that there is greater disability glare and a later recognition of the arrow direction.

The overall simulator test results indicate that amber/white TMA had the highest visibility of work zone, but the lowest visibility of arrow board. The green only TMA had the lowest visibility of work zone, but the highest visibility of arrow board. The tradeoff between work zone visibility and arrow board visibility is an important consideration.

4.2.2 Post-simulator Survey

Post-simulator Survey Methodology

A post-simulator survey was used to collect driver preferences. The 12-question survey consisted of four parts. Part 1 asked participants to rate the daytime light bar configurations. It asked participants to indicate their preferences and rate each daytime configuration with respect to the visibility of work zone vehicles, awareness of work zones, clear recognition of arrow direction and easiness on the eyes. Part 2 asked participants about their preferences for the nighttime configurations. It contained the same questions as Part 1 and one additional question regarding disability glare. Part 3 asked for participants' opinions about simulator fidelity while Part 4 collected demographic information. A 16-question simulator sickness questionnaire (SSQ), which is widely used in diagnosing the simulator sickness severity of participants (Kennedy et al. 1993) was included at the end of the survey. The complete simulator survey is shown in Appendix I.

Participants recruitment was open to the public. Major recruitment methods were printing out information and spread on University of Missouri campus, information circulated among faculty, staff, students, and their friends and family, and recruiting via community connections. An effort was made to balance the proportion in each age group and gender during the recruitment process. However, since the simulator study was conducted in the Columbia metropolitan area, where there are many universities, the majority of participants were between the ages of 16 and 40, and 90% of the participants were from an urban area. Gender was balanced, with 57% male participants and 43% female participants. All participants drove passenger cars as their regular vehicles. Table 4.2.2.1 shows the demographics of the 30 simulator test participants.

Table 4.2.2.1 Demographics information for simulator participants

	Age				Gender		Residency	
	16-25	26-40	41-55	56-70	Male	Female	Urban	Rural
Count	12	13	4	1	17	13	27	3
Percentage	40%	43.3%	13.3%	3.3%	56.7%	43.3%	90%	10%

Post-simulator Survey Results

The following sections describe the post-simulator survey results. Additional comments from the survey are provided in Appendix J.

Part 1: Daytime Evaluation

Part 1 of the survey asked the participants to rank the light bar color options on the TMA for daytime on a scale of 1-5, with 5 representing "like it very much" and 1 representing "dislike it very much". The results from this question are summarized in Table 4.2.2.2. The green/amber scored the highest with an average score of 3.72 while the green only scored the lowest with an average score of 2.97. The amber/white ranked second and the green/white ranked third with scores of 3.48 and 3.14, respectively. The score of 3 represents a neutral opinion, so all configurations were at or above neutral.

Table 4.2.2.2 Daytime TMA configurations preference

	Average score	**Rank**
Amber/white	3.48/5	2
Green only	2.97/5	4
Green/amber	3.72/5	1
Green/white	3.14/5	3

Based on the preference results, a Friedman rank test was carried out to see if there were differences in preference based on TMA configurations. As shown in Table 4.2.2.3, although the average scores were different in each configuration, the non-parametric test shows that rank differences among TMA configurations were not statistically significant.

Table 4.2.2.3 Friedman test: Rank vs. TMA configurations blocked by participants

S = 6.03 DF = 3 P = 0.110 (adjusted for ties)			
TMA Configurations	N	Est. Median	Sum of Ranks
Amber/white	29	2.1	69.5
Green only	29	3.1	82.5
Green/amber	29	1.9	61
Green/white	29	2.9	77

After the preference question, the participants were asked to score different attributes of the various TMA configurations during the daytime on a scale of 1-10, with 1 representing the "lowest" and 10 representing the "highest". These attributes included visibility of work zone vehicles, awareness of work zones, clear recognition of arrow direction, and easiness on the eyes.

Table 4.2.2.4 shows the results of ratings. The amber/white, as the baseline, scored the highest for visibility and awareness of work zones. However, the amber/white TMA scored the lowest in the clear recognition of the arrow direction and easy on the eyes attributes. The green only TMA scored lowest in the visibility and awareness of work zones and scored the highest in the clear recognition of the arrow direction and easy on the eyes attributes. The green/amber and the green/white TMAs scored second and third in all categories respectively. The confidence level for the range of differences was set to be 95 percent in this study based on the Mann-Whitney test. The results show that the amber/white TMA only had a significantly lower score in one category, easy on the eyes, with scoring on average being 5.55, while the green only and green/amber TMAs scored on average 8.62 and 8.07, respectively.

Table 4.2.2.4 Ratings for daytime TMA configurations

	Visibility of work zone vehicles			
	Mean	**Median**	**Diff.**	**Confidence Level**
Amber/white	8.3	9.0	Baseline	
Green only	7.2	8.0	-1.1	92.8%***
Green/amber	8.0	8.0	-0.3	50.6%
Green/white	7.7	8.0	-0.6	78.2%
	Awareness of work zones			
	Mean	**Median**	**Diff.**	**Confidence Level**
Amber/white	8.9	10.0	Baseline	
Green only	7.8	8.0	-1.1	95.0%**
Green/amber	8.5	9.0	-0.4	67.4%
Green/white	8.1	8.0	-0.8	86.8%
	Clear recognition of arrow direction			
	Mean	**Median**	**Diff.**	**Confidence Level**
Amber/white	7.1	7.0	Baseline	
Green only	8.0	8.0	0.9	85.8%
Green/amber	8.0	8.0	0.9	82.6%
Green/white	7.1	7.0	0.0	5.0%
	Easy on eyes			
	Mean	**Median**	**Diff.**	**Confidence Level**
Amber/white	5.6	6.0	Baseline	
Green only	8.6	9.0	2.6	99.9% *
Green/amber	8.1	8.0	2.5	99.9% *
Green/white	6.1	6.0	0.6	51.6%

* indicates significance at 99% confidence level.
** indicates significance at 95% confidence level.
*** indicates significance at 90% confidence level.

Part 2: Nighttime Evaluation

Part 2 of the survey asked the participants to rank the light bar color option on the TMA for nighttime using the same scale of Part 1. The results from this question are summarized in Table 4.2.2.5. The green/amber scored the highest with an average score of 3.59 while the amber/white scored the lowest with an average score of 2.93. The green/white ranked second and the green only ranked third with scores of 3.24 and 3.21, respectively.

Table 4.2.2.5 Nighttime TMA configurations preference ranking

	Average Score	Ranking
Amber/white	2.93/5	4
Green only	3.21/5	3
Green/amber	3.59/5	1
Green/white	3.24/5	2

Another Friedman rank test was conducted, and the test showed that the rank versus TMA configurations by participants were not significantly different from all four TMA configurations. The results are shown below in Table 4.2.2.6.

Table 4.2.2.6 Friedman test: Rank vs. TMA configurations by participants

S = 5.63 DF = 3 P = 0.131 (adjusted for ties)			
TMA Configurations	**N**	**Est. Median**	**Sum of Ranks**
Amber/white	29	3.3	83.5
Green only	29	2.4	74.0
Green/amber	29	1.9	62.0
Green/white	29	2.6	70.5

The participants were asked whether the lights on the TMA were too bright or not, and 66% of participants agreed that the lights of the amber/white were too bright. Meanwhile, only five participants thought the lights of green only were too bright, which is the lowest among all four configurations. The results are shown below in Table 4.2.2.7.

Table 4.2.2.7 Light intensity

Lights too bright?	Agree	Neutral	Disagree
Amber/white	19 (66%)	5 (17%)	5 (17%)
Green only	5 (17%)	3 (10%)	21 (72%)
Green/amber	8 (28%)	11 (38%)	10 (34%)
Green/white	10 (34%)	14 (48%)	5 (17%)

The questions regarding the rating of attributes were then asked using the same format as the daytime part of the survey. The average and median scores for each of the four attributes are shown in Table 4.2.2.8. The amber/white scored the highest in the visibility and awareness of work zones. However, the amber/white also scored the lowest in the clear recognition of arrow direction and easy on the eyes attributes, which showed the same trends for both the daytime and nighttime parts. The green only configuration scored the highest in the clear recognition of arrow direction and easy on the eyes attributes, third in visibility of work zone vehicles and last in awareness of work zones. The green/amber scored second in all four attributes. Finally, the green/white scored last in visibility of the work zone, and third in awareness of work zones, clear recognition of arrow direction, and easy on the eyes.

The Mann-Whitney test shows that the green only TMA had a significantly higher score on clear recognition of arrow direction than the amber/white, as amber/white scored a mean of 6.47 while the green only TMA scored an average of 7.97. For the easy on eyes attribute, the amber/white was significantly lower than all three other TMAs, scoring an average of 4.83. The green only, green/amber, and green/white scored average values of 8.10, 7.50, and 6.20, respectively.

Table 4.2.2.8 Ratings for nighttime TMA configurations

	Visibility of work zone vehicles			
	Mean	**Median**	**Diff.**	**Confidence Level**
Amber/white	8.7	10.0	Baseline	
Green only	8.1	9.0	-0.5	76.5%
Green/amber	8.6	9.5	0.0	44.6%
Green/white	8.0	9.0	-0.7	87.7%
	Awareness of work zones			
	Mean	**Median**	**Diff.**	**Confidence Level**
Amber/white	9.1	10.0	Baseline	
Green only	8.2	8.5	-0.9	93.8%***
Green/amber	8.7	9.5	-0.4	82.4%
Green/white	8.4	9.0	-0.7	88.2%
	Clear recognition of arrow direction			
	Mean	**Median**	**Diff.**	**Confidence Level**
Amber/white	6.5	6.5	Baseline	
Green only	8.0	8.5	1.5	98.3%**
Green/amber	7.6	8.0	1.1	93.7%***
Green/white	7.0	8.0	0.5	56.2%
	Easy on eyes			
	Mean	**Median**	**Diff.**	**Confidence Level**
Amber/white	4.8	5.0	Baseline	
Green only	8.1	9.0	3.3	99.9% *
Green/amber	7.5	8.0	2.7	99.9% *
Green/white	6.2	6.0	1.4	97.2% **

* indicates significance at 99% confidence level.
** indicates significance at 95% confidence level.
*** indicates significance at 90% confidence level.

Part 3: Simulator fidelity

In Part 3 of the survey, questions were designed to evaluate the fidelity of the simulator. Among the participants, 83.3% of them agreed that they felt like they were driving on the highway and that they could drive freely. Only 10% of participants disagreed with both the fidelity and feeling of driving freely in the simulator. The results are shown in Table 4.2.2.9.

Table 4.2.2.9 Simulator fidelity

	Fidelity of highway			Drive freely		
	Count	Percentage		Count	Percentage	
Strongly Agree	9	30%	83.3%	9	30%	83.3%
Agree	16	53.3%		16	53.3%	
Neutral	2	6.7%	6.7%	2	6.7%	6.7%
Disagree	2	6.7%	10%	3	10%	10%
Strongly Disagree	1	3.3%		0	0%	

Simulator sickness questionnaire (SSQ) results

The post-simulator survey was followed by a 16-question SSQ, to assess motion sickness of participants after the simulator experiment. The results are shown in Table 4.2.2.10. Most participants felt no or slight discomfort, and only five out of 30 participants had moderate or severe symptoms.

Table 4.2.2.10 Summary of SSQ Results

	General discomfort		Fatigue		Headache		Eye strain	
	count	pct.	count	pct.	count	pct.	count	pct.
None	24	80%	27	90%	24	80%	16	53.3%
Slight	6	20%	1	3.3%	6	20%	11	36.7%
Moderate	0	0%	2	6.7%	0	0%	3	10%
Severe	0	0%	0	0%	0	0%	0	0%
	Difficulty focusing		**Salivation increasing**		**Sweating**		**Nausea**	
	count	pct.	count	pct.	count	pct.	count	pct.
None	26	86.7%	29	96.7%	28	93.3%	25	83.3%
Slight	4	13.3%	1	3.3%	2	6.7%	4	13.3%
Moderate	0	0%	0	0%	0	0%	0	0%
Severe	0	0%	0	0%	0	0%	1	3.3%
	Difficulty concentrating		**Fullness of the Head**		**Blurred vision**		**Dizziness with eyes open**	
	count	pct.	count	pct.	count	pct.	count	pct.
None	29	96.7%	22	73.3%	28	93.3%	26	86.7%
Slight	1	3.3%	8	26.7%	2	6.7%	3	10%
Moderate	0	0%	0	0%	0	0%	1	3.3%
Severe	0	0%	0	0%	0	0%	0	0%
	Dizziness with eye closed		**Vertigo**		**Stomach awareness**		**Burping**	
	count	pct.	count	pct.	count	pct.	count	pct.
None	28	93.3%	29	96.7%	29	96.7%	30	100%
Slight	2	6.7%	1	3.3%	1	3.3%	0	0%
Moderate	0	0%	0	0%	1	3.3%	0	0%
Severe	0	0%	0	0%	0	0%	0	0%

Summary of post-simulator survey results

The overall post-simulator survey results indicate that amber/white TMA was too bright and not easy on eyes, and the visibility of the green only TMA was too low. On the other hand, the results indicate that the green only TMA allowed clear recognition of the arrow direction and was easy on eyes. Green/amber scored in between amber/white and green only and was preferred in both daytime and nighttime scenarios.

4.2.3 Field Test

Two major tasks for this project were to conduct simulator and field studies to evaluate the performance of different TMA light configurations. After the simulator tests, a field test was performed to collect data on field performance for two of the configurations. The field studies involved usage of Lidar and video recording to ensure accuracy of measure.

Field Test Set up

The field test examined the amber/white TMA and green only TMA deploying the TMAs in the same way as in a regular mobile work zone. The test was conducted on a section of US 50, a four-lane freeway, near Kansas City, Missouri. A TMA started eastbound near the intersection of MO-291 and US 50 at the speed of 10 to 20 mph and turned around at the Buggy Stop at the intersection of US 50 and MO 58, near Centerview, Missouri, as shown in Figure 4.2.3.1. The data collection took place on two days: December 19th, 2017, and December 20th, 2017. The green only TMA was tested on the first day, and the amber/white TMA was tested on the second day. Both TMAs were tested for four hours during the daytime and four hours during the nighttime. On the first day, the green only TMA was utilized as the rear advance TMA. However, due to an equipment availability issue, there was only one green only TMA available, so the green only TMA was followed by an amber/white shoulder TMA. On the second day, two amber/white TMAs were utilized, one as the rear advanced TMA and the other as the shoulder TMA.

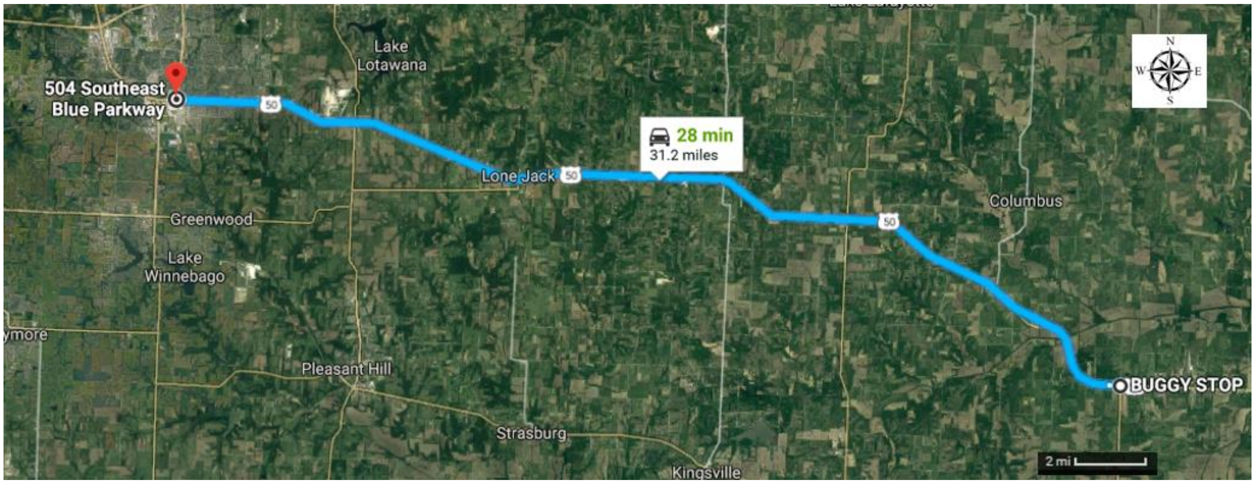


Figure 4.2.3.1 Route travelled for data collection on US 50 (Google maps 2017)

In order to capture the distance of vehicles from the rear advance TMA, a LiDAR detector was utilized. The LiDAR used was an active infrared detector, and it rated as eye safe. It was mounted on the TMA with a LED display. A camera was mounted on the TMA to record traffic behavior and LED readings for post video image processing. This video perspective was useful since the traffic conditions surrounding the LiDAR readings were captured. The MoDOT custom-mounted LiDAR is shown in Figure 4.2.3.2, and an example of LiDAR capturing vehicle distance from rear advance TMA is presented in Figure 4.2.3.3. Although the decimal dot was not shown on the LED monitor, data was recorded with two decimal points. In the example of Figure 4.2.3.3, the distance of the vehicle was 142.47 feet from the rear end of the rear advance TMA.

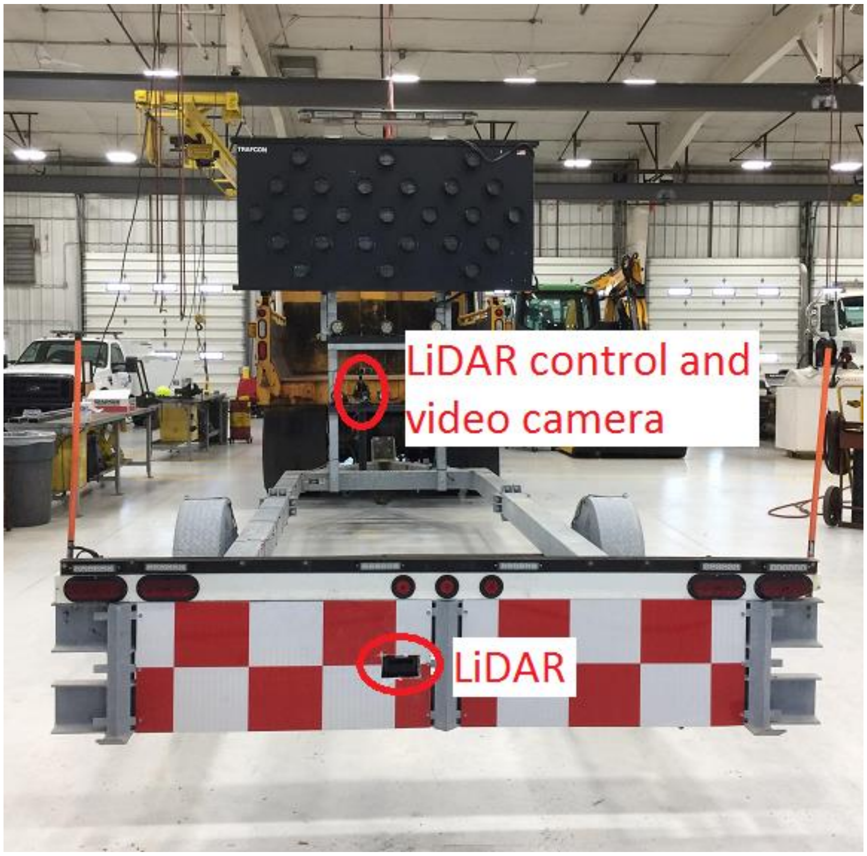


Figure 4.2.3.2 MoDOT custom-mounted TMA

Figure 4.2.3.3 Example of LiDAR capturing distance from vehicle to rear advance TMA

Video Image Processing Methodology

After the field study was completed, videos recorded from the field were reviewed by the team to extract data. Video recording started when the TMAs departed the garage, but the data was not valid until TMAs started to operate as a mobile work zone. Within the valid data collecting time period, 4,966 vehicles passed through the mobile work zone, and 2,589 of them were platoon leaders. There were 2,756 instances in which the LiDAR captured distance measurements to various objects behind it, such as vehicles in the travel lanes, roadside barriers, or the shoulder TMA. The general information from the data collection is shown in Table 4.2.3.1.

Table 4.2.3.1 General information

	19th daytime (green only)	19th nighttime (green only)	20th daytime (amber/white)	20th nighttime (amber/white)
Total vehicles counted	1079	787	1918	1182
Total vehicles shown on right lane within field of view	58	4	78	12
Vehicles merged after passing shoulder TMA	2	2	1	3

The speed of vehicles when they passed the rear advance TMA was the only measurement of effectiveness processed in the field study. Vehicle speeds were derived from videos using photogrammetry. This technique allows the derivation of speeds of any vehicles when driving past a TMA, in contrast to speed radars which typically track vehicles only within a limited distance or angle. When a vehicle approached the rear advance TMA and passed a beginning dash line selected by the analyst, the time stamp was recorded as T_1, as shown in Figure 4.2.3.4. Then the analyst selected an end dash line (say the fifth dash), and when the vehicle passed the end dash line, the time stamp was recorded, as T_2, as shown in Figure 4.2.3.5. The number of dash lines used varied from two to five, depending on visibility due to geometry and nighttime lighting conditions. The distance travelled was 40 feet multiplied by the number of dashes traveled, and the travel time was T_2 minus T_1. From the travel distance and travel time, the vehicle speed was calculated. This technique was validated by computing the speed several times using different number of dash lines; the results were consistently the same regardless of the number of dashes used.

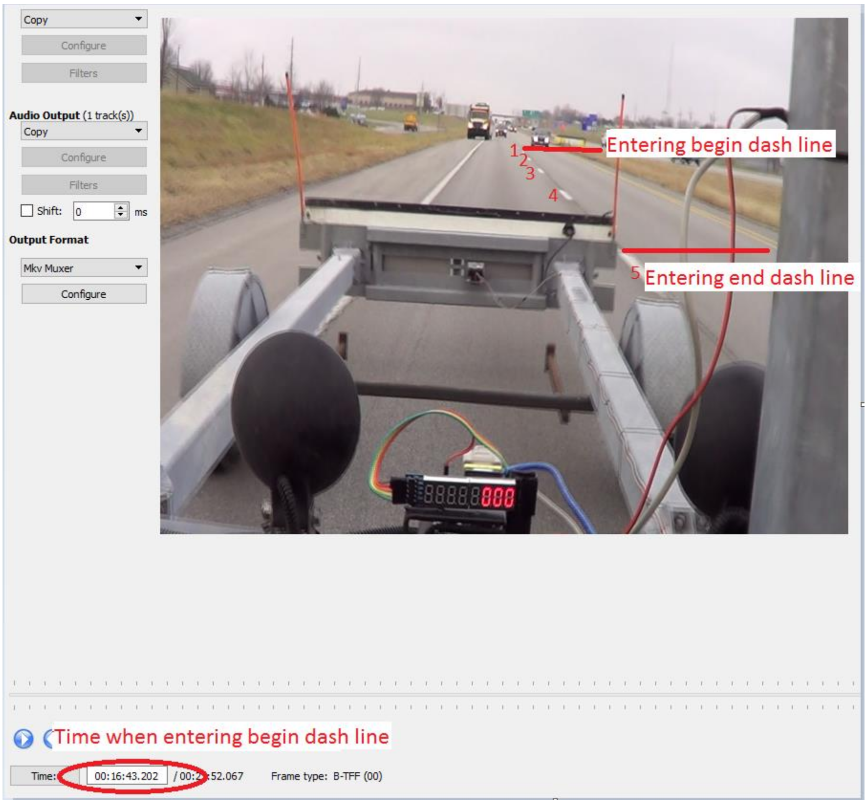


Figure 4.2.3.4 Time stamp one: vehicle entering begin dash line

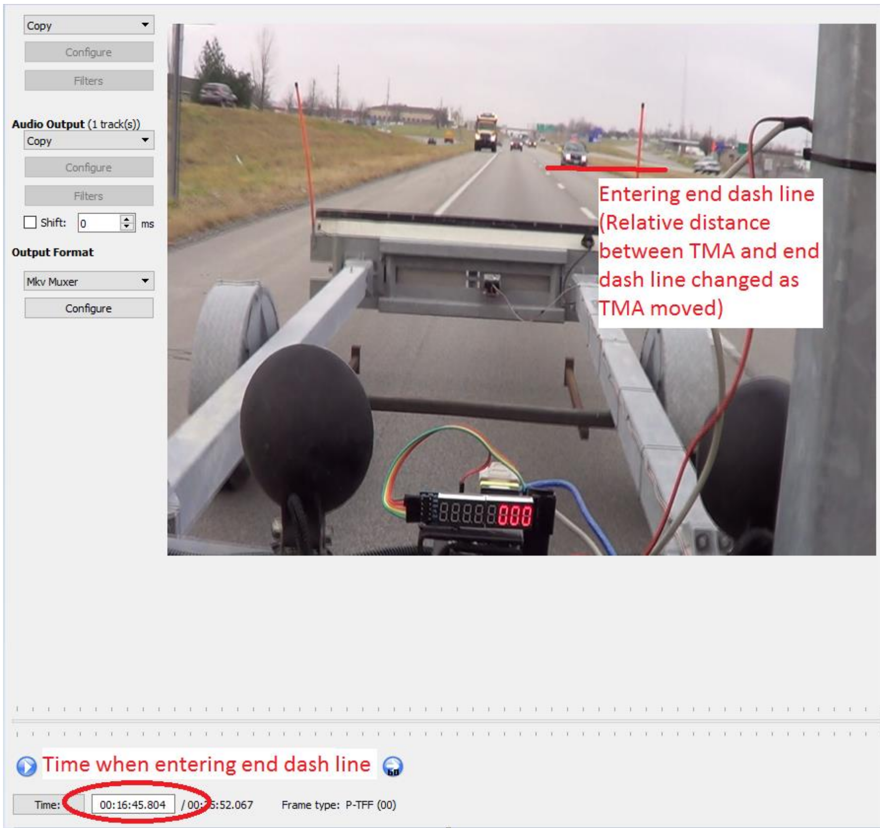


Figure 4.2.3.5 Time stamp two: vehicle entering end dash line

Besides the impact of different light bar configurations, the speed of vehicles may be affected by many other factors, such as the speed of a platoon leader if in a platoon, the speed of TMA, entering or exiting ramp, curves, or commercial vehicles and buses. To limit the impacts of other factors, data for vehicles entering and exiting the highway was discarded, as they were not reacting to the TMA but to entering/exiting maneuvers. Only the platoon leader speeds were captured to eliminate the influence of vehicle platooning. The speed of the rear advance TMA was sampled every five minutes and combined with vehicle passing speed to see how vehicles reacted to different TMA speeds. A total of 186 TMA speeds were sampled.

Field Results

During the two-day field study, the travel speeds of the TMAs were not consistent due to different drivers and other factors. The speed of the green only TMA was consistent during daytime, but the speed of the amber/white TMA was significantly different from morning and afternoon. As TMA speed could influence vehicle passing speed, a statistical test was run for the speed data, with amber/white as the baseline. As shown in Table 4.2.3.2, the amber/white TMA travel speed in the afternoon was significantly lower than in the morning, and the speed of leading vehicles in the afternoon was significantly lower than in the morning. This is potential evidence that lower TMA speeds lead to lower vehicle passing speeds.

Table 4.2.3.2 Daytime speeds for amber/white (mph)

	Daytime speed: amber/white (mph)			
	TMA Speed		**Leader speed**	
	Morning	**Afternoon**	**Morning**	**Afternoon**
Count	13	35	186	566
Mean	20.3	9.8	64.5	61.9
Stdev.	1.1	2.9	6.0	6.9
Confidence level	99.9%*		99.9%*	
Cohen's d	4.76		0.40	
Effect size	Large		Small	

* indicates significance at 99%.

The results of the statistical analysis indicate that the TMA speeds may impact vehicle passing speeds. Therefore, the field data results are presented along with different TMA speeds. The overall mean daytime speeds are shown in Table 4.2.3.3. The amber/white TMA speed was significantly lower than the green only TMA speed, and the leading vehicle passing speed for amber/white TMA was slightly higher than for the green only TMA. It could be implied that if TMAs were travelling at the same speed, the leading vehicle passing speed for green only TMA would be much lower. In terms of impacts on vehicle speeds, the daytime performance of green only TMA was more desirable.

Table 4.2.3.3 Mean daytime speed (mph)

	Mean daytime speed (mph)			
	TMA Speed		Leader speed	
	green only	amber/white	green only	amber/white
Count	45	48	702	752
Mean	19.0	12.6	62.5	62.6
Stdev.	1.6	5.4	6.1	6.7
Confidence level	99.9%*		52.0%	
Cohen's d	1.67		n/a	
Effect size	Large		n/a	

* indicates significance at 99%.

In order to test if drivers behaved similarly under comparable TMA speeds, another statistical test was performed to compare the green only daytime data and amber/white morning data. The average morning amber/white TMA speed was 20.3 mph which was close to the green only TMA speed of 19 mph. The results are shown in Table 4.2.3.4. Although the speed of leading vehicles for green only TMA was slower, the speed of green only TMA was slightly slower than amber/white as well.

Table 4.2.3.4 Mean speed: green only daytime vs. amber/white morning (mph)

	Mean speed: green only daytime vs. amber/white morning (mph)			
	TMA Speed (mph)		Leader speed	
	green only daytime	amber/white morning	green only daytime	amber/white morning
Count	45	13	702	186
Mean	19.0	20.3	62.5	64.5
Stdev.	1.6	1.1	6.1	6.0
Confidence level	99.8%*		99.9%*	
Cohen's d	0.90		0.32	
Effect size	Large		Small	

* indicates significance at 99%.

During nighttime, even though the green only TMA travelled at a significantly higher speed than amber/white TMA, the vehicle passing speed for green only TMA was slightly lower, as shown in Table 4.2.3.5. The results show that green only TMA performed better than amber/white TMA during nighttime.

Table 4.2.3.5 Mean nighttime speed (mph)

	Mean nighttime speed (mph)			
	TMA speed		Leader speed	
	green only	amber/white	green only	amber/white
Count	45	48	504	631
Mean	22.2	12.6	52.1	52.9
Stdev.	2.7	1.2	8.4	7.9
Confidence level	100.0%*		94.6%***	
Cohen's d	4.60		0.10	
Effect size	Large		Small	

* indicates significance at 99%.
*** indicates significance at 90%.

For the same TMA configuration, driver behavior in daytime and nighttime was different. As shown in Table 4.2.3.6, even though the green only TMA travelled faster during nighttime, the vehicle passing speed was significantly slower during nighttime.

Table 4.2.3.6 Green only speed: daytime vs. nighttime (mph)

	Mean speed for green only (mph)			
	TMA Speed		Leader speed	
	Daytime	Nighttime	Daytime	Nighttime
Count	45	45	752	504
Mean	19.0	22.2	62.6	52.1
Stdev.	1.6	2.7	6.7	8.4
Confidence level	99.9%*		99.9%*	
Cohen's d	1.43		1.36	
Effect size	Large		Large	

* indicates significance at 99%.

The mean speeds of amber/white were similar for daytime and nighttime. As shown in Table 4.2.3.7, under the same TMA speeds, drivers tended to slow down during nighttime, and the passing speed at nighttime was significantly slower than daytime by almost 10 mph.

Table 4.2.3.7 Amber/white speed: daytime vs. nighttime (mph)

	Mean speed for amber/white (mph)			
	TMA Speed		Leader speed	
	Daytime	Nighttime	Daytime	Nighttime
Count	48	48	752	631
Mean	12.6	12.6	62.6	52.9
Stdev.	5.4	1.2	6.7	7.9
Confidence level	50.4%		100.0%	
Cohen's d	n/a		1.32	
Effect size	n/a		Large	

* indicates significance at 99%.

Luminance Measurement

TMA light bar luminance was measured in order to study the light intensity of different TMA configurations. A luminance meter was utilized for measurement. Each light bar was measured three times to obtain the maximum and minimum light intensity. The mean of three readings is shown in Table 4.2.3.8. The data shows that the intensity of green light was higher than amber and white lights.

Table 4.2.3.8 Luminance measurement of TMA light bars (cd/m^2)

	Max	Min
Green (Undimmed)	6800	28
Amber (Undimmed)	7033	1329
White (Undimmed)	6333	551

It was difficult to measure the light intensity due to different flash patterns, because under different flash patterns, the intensities of lights were not consistent. Also, the dim/undim switch control was not available for amber/white TMA. There was a sensor of day light intensity on amber/white TMA, and the light bars were controlled by the day light sensor. When it was dark outside, the amber/white TMA was dimmed automatically. Also, there were some issues related to the calibration of the NIT gun. For future studies, it would be helpful to have greater laboratory control over the light bars.

Summary and Discussion of Field Study Results

The results for vehicle passing speeds showed multiple findings:

- Lower TMA speed led to lower vehicle passing speeds.
- Vehicles slowed down during nighttime.
- The green only TMA performed better than the amber/white TMA, in both daytime and nighttime. This result could be due to various reasons. It could be that green color caught people's eyes better, or it could be the novelty effect of the green only TMA.

The field study results should be viewed with some considerations in mind. As the green color was deployed for the first time on work vehicles in Missouri, the novelty effect of green light TMAs may have impacted the results. Once drivers become familiar with green light TMAs, the performance of the configuration may be different. The results may vary in a different state, and the novelty effect of green lights should be examined in a study of longer duration. The field test was conducted during winter, and the green color stood out from the background of road. It is unclear if the effect of green light will be the same during spring and summer because of the lack of contrast due to green foliage.

Due to the limited availability of green only TMAs, the shoulder TMA was amber/white instead of green during the green only data collection. Since drivers passed the shoulder TMA first, they may have reacted to the amber/white shoulder TMA instead of the green only rear advance TMA. Because the TMA speeds were not consistent between the baseline and green only TMA, it was challenging to determine the magnitude of the effects caused by TMA speeds and light color configurations.

The field study attempted to record merge distance as one critical measurement from the perspective of collision prevention. However, most of the vehicles did not merge near TMAs, and only 8 out of 4,966 vehicles merged to the passing lane after bypassing the shoulder TMA; most merged upstream of the shoulder TMA.

4.2.4 Conclusions

The simulator tests and post-simulator survey indicated that the amber/white TMA had the highest work zone visibility but lowest arrow recognition, and the green only TMA had the reverse effect. An inverse relationship between visibility/awareness of work zone and arrow board recognition/easy on eyes was derived from the results. This is consistent with literature, such as the NCHRP study on warning lights on roadway maintenance vehicles (Gibbons 2008). The green/amber TMA light configuration performed roughly between the two aforementioned configurations.

The field test results showed that lower TMA speeds led to lower vehicle passing speeds. In addition, the green light TMA performed better than the amber/white TMA, as the overall passing speed for the green only TMA was significantly lower. It was also found that drivers slowed down significantly during nighttime. The results from simulator study and field tests were complementary to each other, and all four configurations appear to be viable although none is clearly superior.

The study made several original and significant contributions to the body of existing knowledge regarding the use of green lights on work vehicles. This was the first study to quantitatively assess the use of green lights on work vehicles in the United States. In addition, the Missouri testing was the first on TMAs, and the results of this project are going to assist in making an informed decision on whether or not to implement green lights. Other DOT implementations have currently been limited to snow removal vehicles. The study included both simulator and field testing. The driving simulator allowed for multiple configurations to be tested efficiently in a controlled environment. Implementation of an eye tracker in the disability glare test provided for the tracking of human eye movement to determine if a specific object was recognized. The field study yielded data regarding vehicle speeds in a real-world implementation of green lights on TMAs.

The results of this project could be enhanced in many ways. Deploying a green only shoulder TMA to accompany the green only rear advance TMA would help to ensure that only the green light configuration was seen. Other configurations, such as green/amber could be tested in the field. The TMAs should be driven at the same speed to control for the effects of TMA speed on traffic speed. An additional study of longer duration could be undertaken to assess any possible novelty effects. A field evaluation during spring or summer could help to determine if there are any seasonal effects due to foliage color.

The study represented an effort to try and improve work zone visibility. Enhancing mobile work zone visibility with low cost countermeasures can help provide warning to drivers of the work zone and improve safety for both drivers and the work crew. However, MoDOT decided not to implement green light TMAs as the results of this study indicated that all four options were viable, and a change in state law would be required to allow use of green lights in Missouri.

Mobile work zone safety can be improved by other complementary measures, such as 1) better public education on texting and driving, especially near work zones; 2) exploration of other warning systems such as audio alerts in addition to visual cues; 3) working with connected autonomous vehicle industry to develop new technologies that could automatically recognize mobile work zones in the future.

The study of green lights on TMAs could be transferred to other states, and the result and conclusion may be different from this study. The field study was conducted in winter in

Missouri, when there was no green tree, grass, or leaf, and the contrast was high. If tested during other seasons, or in states with warmer winter, green light may not stand out from the background. Another concern about green light TMA test conducting in other states is, if the traffic volume is high, then the vehicles would follow platoon leader, and the reactions would be between platoon followers and platoon leader, instead of vehicles and green light TMA. As green is not a commonly used color in work zone alarms, law change or permission may be required to test or deploy green light TMAs for mobile work zones in other states.

4.3 AUTONOMOUS TRUCK PLATOON IN WORK ZONE

Low level of automation is expected to be widely deployed in the near future. Reduced fuel consumption, higher efficiency, and lower freight cost are some of the potential benefits. However, public acceptance of truck platoon is a challenge, as fatal risk of crashes involving trucks, especially near work zones, is much higher. Smart Work Zone Deployment Initiative (SWZDI) funded a project to investigate autonomous /connected vehicles in work zones. The objective of this project is to examine truck platoon and human driven vehicles interaction near work zones, and the goal is to figure out ways to improve smart work zone safety.

This is a simulator study only project, as autonomous vehicles are not ready for real life common utilization yet. The objectives of the project are: to analyze driver behavior, to explore policy issues, and formulate and provide recommendation for DOTs.

The results of this study are going to provide DOTs with possible policy recommendations, including signage/message to be displayed on the back of autonomous trucks, public education change, adding extra content for driver training, and incorporating driver preference obtained from survey in policy.

There are five tasks in research plan. Task 1 is to work with technical advisory committed to get the project plan settled. The project kick-off meeting was held on June 25, 2019. Task 2 is literature review. Current literature reviewed is categorized into four groups: 1) truck crashes near work zones, 2) truck work zone safety improvement, 3) truck platoon, and 4) autonomous vehicle in work zones. However, there is not much literature currently that is related to autonomous truck platoon in work zones, which means this project is original and significant. Task 3 is simulator study. As the simulator study involves human subjects participating simulator trials and take post-simulator survey, the project has to be reviewed and approved by IRB before participant recruitment. IRB has approved this project as exempted. The current stage of simulator study is scenario development and survey design. Task 4 is performance measure and data analysis, and it is going to be processed right after participants finish up simulator trials. Task 5 is report preparation and review, and all deliverables are originally scheduled to be delivered by July 2020. However, due to the pandemic of Coronavirus (COVID-19), ZouSim lab has been lockdown since March 23rd, 2020. Until the defense of this dissertation (May 4th 2020), there is not notification of lab reopen.

4.3.1 Simulator Study

This simulator study utilizes a federated simulator system, with a car simulator and a truck simulator connected and interact with each other via a virtual server wirelessly, as shown in Figure 4.3.1.1. The car simulator serves as the host, also called as local client, and the truck simulator is the remote client.

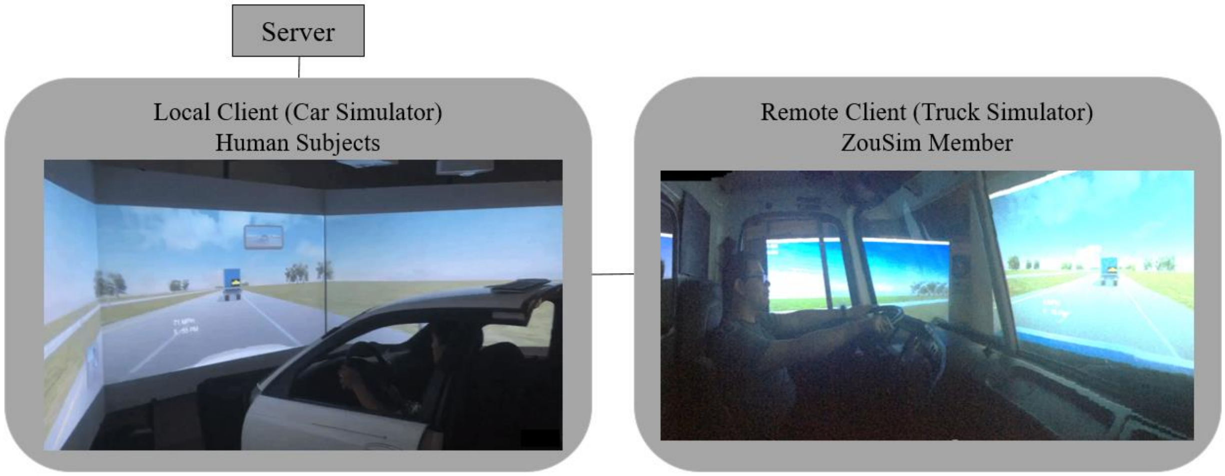


Figure 4.3.1.1 Federated Simulator System

The setup of work zone is on a two-way four-lane highway, with shoulder closure. An autonomous truck platoon, with leading truck driven by ZouSim team member, is leading the traffic, following by a participant driven sedan. The sedan needs to exit a ramp after bypassing the work zone.

Three factors are tested in the study. The first factor is education. Participants start without education on the meaning of signage/message displayed on the back of truck. After completing four trials, they take a half-way break, and the hosts educate participants about the meaning of display and inform them about truck platoon. The second factor is truck platoon display. The baseline of this factor is no display, comparing with scenarios where there is a signage/message on the back of the truck to indicate that trucks are travelling in a platoon. The third factor is number of trucks in the platoon. A platoon with two trucks is compared with a platoon with four trucks. Scenarios are listed in Table 4.3.1.1. The scenarios examining display and number of trucks are randomized in order to eliminate sequence bias.

Table 4.3.1.1 Proposed Scenarios

Scenario	Education	Number of Trucks	Sign	Order
1	No	2	No	Randomized
2	No	4	No	
3	No	2	Truck Platoon	Randomized
4	No	4	Truck Platoon	
5	No	2	2 Trucks	
6	No	4	4Trucks	
7	Yes	2	No	Randomized
8	Yes	4	No	
9	Yes	2	Truck Platoon	Randomized
10	Yes	4	Truck Platoon	
11	Yes	2	2 Trucks	
12	Yes	4	4Trucks	

The design of signage is shown in figures below. Figure 4.3.1.2 is the appearance of truck platoon when no signage is displayed on the back of trucks. Figure 4.3.1.3 shows number of trucks in the platoon displayed on the back of trucks. Figure 4.3.1.4 displayed “Truck Platoon”. There are three reasons why the display monitor is in the middle right side: 1) eye tracker is utilized in the experiment, a small monitor helps researcher to determine if participants look at specific place; 2) work zone closes the right lane, truck platoon starts at left lane and car starts at right lane, putting the sign on the right side could help participants see the contents clearer; 3) consideration of manufacture cost.

Figure 4.3.1.2 No Signage

a) Signage shows “2 Trucks” b) Signage shows “4 Trucks”

Figure 4.3.1.3 Signage Shows Number of Trucks in the Platoon

Figure 4.3.1.4 Signage Shows “Truck Platoon”

There could be three types of potential driver behavior: car driver slows down and follows platoon to pass the work zone and then exit the ramp; car driver speeds up and bypass the platoon when encountered work zone, and merge back to right lane to exit; car driver squeezes in between the trucks. More uniform performance is expected after education.

Seven measures of effectiveness (MOEs) are captured from the trials. MOE 1 is driver behavior. MOE 2 is the distance between work zone and car when it merges (ft.). MOE 3 is the speed of car when it merges (mph). MOE 4 is the distance between car and back of the last ruck in the platoon when car merges (ft.), if car follows the truck platoon to pass. MOE 5 is the distance between car and head of the leading ruck in the platoon when car merges (ft.), if car speeds up and bypasses the truck platoon. MOE 6 is record of brake of the car. MOE 7 is record of blinker used by car.

4.3.2 Post-simulator Survey

A post-simulator survey is administrated to collect participants' demographic information, and obtain their preference of factors. The survey also asked for participants' opinions about the effectiveness of education and how clear and effective the signage/message conveys information. Simulator fidelity is examined, followed by an SSQ (Kennedy et al 1993). The draft of survey is attached as Appendix K.

CHAPTER 5: APPLICATION OF STATISTICAL ANALYSIS IN TRANSPORTATION ENGINEERING

5.1 FACTOR ANALYSIS ON SYNERGY OF FIELD AND SIMULATOR STUDIES

Exploratory Factor Analysis (EFA) is a statistical approach to find out the correlation of variables and determine the underlaying structure in a dataset (Ford et al. 1986). Generally, EFA is a prior step of structural equation modeling, such as Confirmatory Factor Analysis (CFA).

Confirmatory Factor Analysis (CFA) is a type of structural equation modeling that deals with the relationship between observation and latent variables (Brown 2014). CFA is most commonly used in social research (Kline 2005). The objective of CFA is to test whether data fits a hypothesized model (not able to reject null hypothesis is desired), which was built based on practical experiments or understanding.

One critical assumption for EFA and CFA is that all variables to be structured have to be non-nominal, which means that some categorical or binary data would not be able to be processed by EFA. When it is applied to Likert scale variables, the scale needs to be converted to numerical values and input as continuous variables. Another key point is that it is very rare in EFA and CFA that objective variable in included in the structure, because the structure contains only reflective latent variables, from which latent factors could be identified, and objective variable is rarely a reflective latent variable. Missing data should either be removed or be estimated.

Data from AFADs field and simulator studies are used as an example to demonstrate Factor Analysis application. The most important purpose presented in this chapter is to explore the new methodology. Due to limited sample size of simulator participants and field survey response, and data were sourced from different persons, the results may not be desirable. The sample size of the dataset was 115, and data information is listed in Table 5.1.0.1.

Table 5.1.0.1 Data Information for Factor Analysis

Variable	Source	Measurement type	Range	Pre-processing method
Brake	Simulator trial	Continuous	68-637	Standardized using z-score
Speed	Simulator trial	Continuous	9-58	Standardized using z-score
Stop	Simulator trial	Continuous	7-247	Standardized using z-score
Understanding	Simulator survey	Binary	0-1	Removed
Clarity	Simulator survey	Ordinal	1-10	Regarded as continuous, standardized using z-score
Visibility	Simulator survey	Ordinal	1-10	Regarded as continuous, standardized using z-score
Safety	Simulator survey	Ordinal	1-10	Regarded as continuous, standardized using z-score
Efficiency	Simulator survey	Ordinal	1-10	Regarded as continuous, standardized using z-score
FieldSpeed	Field test	Continuous	10-40	Standardized using z-score
FieldStop	Field test	Continuous	10-220	Standardized using z-score
FieldClarity	Field survey	Ordinal	1-10	Calculated from effectiveness, regarded as continuous, standardized using z-score
FieldVisibility	Field survey	Ordinal	1-10	Calculated from effectiveness, regarded as continuous, standardized using z-score
FieldSafety	Field survey	Ordinal	1-10	Calculated from effectiveness, regarded as continuous, standardized using z-score
FieldEfficiency	Field survey	Ordinal	1-10	Calculated from effectiveness, regarded as continuous, standardized using z-score

5.1.1 Exploratory Factor Analysis (EFA)

One strength of EFA over CFA is that there is no prior theory that limits the number and specific variables in the structure, which provides more freedom in structuring the models. Another strength of EFA is that it screens out the problematic variables, which is assumed to be within the latent factor but actually not, while CFA only tells that latent factor could not be identified without pointing out the problematic variables. Some statistical software such as SAS, R, and IBM SPSS have built-in functions of EFA.

The application of EFA in transportation engineering could be determining the correlation between MOEs and assign them as groups in a structure. Further process of CFA could be done to verify if the combinations of variables match with existing theory or researcher assumptions.

Modelling Simulator and Post-Simulator Survey Data

Use the data of AFAD simulator first trial and survey data as an example. Variables in this dataset are: participants' understanding of traffic controllers (binary, with zero represents misunderstanding and one as correct understanding), participant rating of AFADs and human flagger (denoted as clarity, visibility, safety, and efficiency, rating from one to 10, as one being the lowest score and 10 being the highest score), first brake distance from work zone (unit: ft., denoted as Brake), speed at 250 ft. location from work zone (unit: mph, denoted as Speed), and full stop distance from work zone (unit: ft., denoted as Stop). Researcher was attempting to uncover the underlaying relationship of these variables and do further process in CFA. The EFA process was conducted in IBM SPSS, with component grouping extracted via Principle Component Analysis (PCA) method and matrix rotation via Varimax with Kaiser Normalization. PCA reduces dimension of datasets, increases interpretability, and minimizes information loss by creating new uncorrelated variables with maximized variance (Jolliffe & Cadima 2016). Varimax rotation simplifies the expression of a particular space, with actual coordinate system unchanged, and only orthogonal basis rotated to align with coordinates. Kaiser Normalization normalizes factor loadings before rotating them, and then denormalizes them after rotation (Kaiser 1958). Through this rotation process, a rotated component matrix is obtained as shown in Table 5.1.1.1.

Table 5.1.1.1 Rotated Component Matrix

	Component		
	1	2	3
Understanding			0.900
Clarity	0.815		
Visibility	0.894		
Safety	0.923		
Efficiency	0.786		
Brake		0.769	
Speed		-0.607	0.618
Stop		0.494	

Table 5.1.1.1 indicates that the ratings should be grouped together for one latent factor, and Brake, Speed, and Stop structure another latent factor, however, the correlation of speed and the other two variables is negative. Understanding should stand alone, however, it is rare to use only one variable as a reflective variable group, and understanding was binary data when it was collected, then it is removed from future CFA.

5.1.2 Confirmatory Factor Analysis (CFA)

1. Modelling Simulator and Post-Simulator Survey Data

Researcher was attempting to identify a latent variable “safety” from the reflective variables. Prior assumptions are: 1) when drivers are able to understand the meaning of the new configurations, the probability of them following instructions is higher, which is safer; 2) higher clarity and visibility of traffic control devices would alert drivers with the existence of work zone earlier, then drivers slow down and prepare to stop, which creates a safer work zone environment; 3) the rating of safety should be consistent with latent safety variable; 4) when efficiency is rated high, that means message is conveyed effectively and drivers would follow instruction better; 5) earlier braking, lower approaching speed, and further full stop distance are regarded as safer features.

<u>Original Model</u>

With the results of EFA in hand, a two-factor CFA model was built. The equation set of this model is written as

$$
\begin{aligned}
\text{Brake} &= \text{lam1 f1} + \text{e1} \\
\text{Speed} &= \text{lam2 f1} + \text{e2} \\
\text{Stop} &= \text{lam3 f1} + \text{e3} \\
\text{Clarity} &= \text{lam4 f2} + \text{e4} \\
\text{Visibility} &= \text{lam5 f2} + \text{e5} \\
\text{Safety} &= \text{lam6 f2} + \text{e6} \\
\text{Efficiency} &= \text{lam7 f2} + \text{e7}
\end{aligned}
$$

with the series of lam representing parameters to be measures for each observed variable, series of e representing random errors with variance. Latent variable Simulator Safety was denoted as f1, latent variable Survey Safety was denoted as f2, and the correlation between these two factors is rho12. The total number of parameters to be estimated is 15, and they are the seven lambdas, seven variances, and correlation rho12. The CFA model was analyzed via AMOS, a CFA software affiliated under IBM SPSS. The path diagram is shown in Figure 5.1.2.1. After removing all missing data, the final sample size was 115.

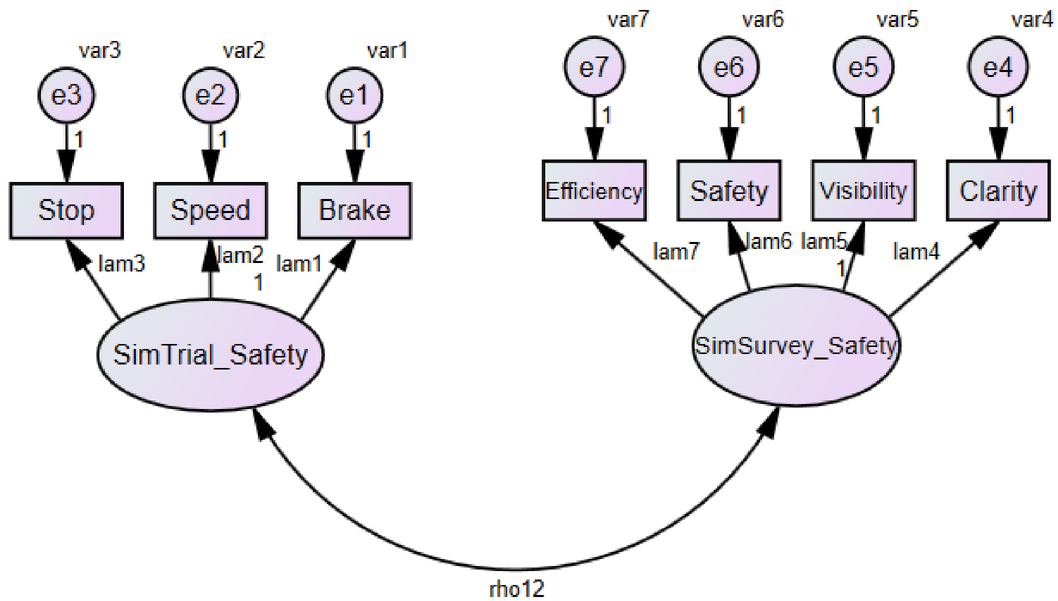


Figure 5.1.2.1 Path Diagram of Two-Factor CFA Model

Model fit statistics measures if the construct of data is consistent with researcher's understanding (Preedy 2010). If the model fit is good, in common situation adjusted goodness of fit index (AGFI) over 0.8 and RMR less than 0.05, then it indicates that data is consistent with factors, and the way how researchers categorize data is legitimate. If the model fit is poor, it implies inconsistency, or high correlation within items under factor.

The model fit results are shown in Table 5.1.2.1. The Chi-square test value is small and the p-value is large, resulting in failing to reject the null hypothesis of consistency, which is desirable in CFA. Although GFI and AGFI seemed to be high, the RMR is very large, which means that the model could be improved.

Table 5.1.2.1 Model Fit of Simulator Safety Model

Chi-Square	P-value	RMR	GFI	AGFI
13.733	0.393	97.461	0.968	0.930

The estimated values of parameters are shown in Table 5.1.2.2. The results indicate consistency with the assumptions made by researcher. Early brake, low speed, and far stop are desirable for safe work zones, as coefficients of brake and stop were positive and coefficient for speed was negative. Clarity, visibility, safety and efficiency are reflective variables of work zone safety because of positive coefficients, and among them, visibility and rating of safety are weighted heavier. The estimate of correlation between simulator safety and survey safety rho12 was 0.538, indicating that the latent safety factors are correlated with each other. However, the correlation was not very strong they were not able to explain the variance of each other well.

Table 5.1.2.2 Summary of Estimates of Simulator CFA Model

Latent Factor	Variable	Estimates		
		Lambda	Variance	Correlation
SimTrial_Safety	Brake	40.638	11686.681	
	Speed	-3.965	109.439	
	Stop	20.499	1627.513	
SimSurvey_Safety	Clarity	1.9	3.107	0.538
	Visibility	2.423	1.605	
	Safety	2.479	0.642	
	Efficiency	1.733	3.134	

There are some shortcomings in the CFA model. The scales of data magnitude were different in two factor groups: survey group were ratings from one to 10, and the simulator data group were real continuous data capturing driving behaviors, with Brake ranging from 68 to 637 ft., Speed ranging from nine to 58 mph, and Stop ranging from seven to 247 ft. One way to get this issue resolved it to standardize the data, or recode the simulator data to fit them in a one to 10 scale, then feed them into CFA. Another shortcoming of this model is the desired correlation between Speed and latent variable safety is naturally negative, which is not in favor of CFA modeling. To get this issue fixed, researcher may recode the data, and assign lower speed to higher score.

Improved Model

As the original model did not perform well due to the scale difference, a new model with all data standardized is built to improve. Data standardization was complete via built-in function of IBM SPSS, using the Z-score method, of which data was given a common standard with mean of zero and a standard deviation of one. The new variables were created in the same dataset with the names adding a “Z” as prefix. The model fit summary is shown in Table 5.1.2.3, with small Chi-square, large P-value, small RMR, and high AGFI, which indicate that this model is a good fit.

Table 5.1.2.3 Model Fit of Improved Simulator Safety Model

Chi-Square	P-value	RMR	GFI	AGFI
3.733	0.393	0.043	0.968	0.930

The estimates are listed in Table 5.1.2.4. From the sign of the estimated lambdas, it could be implied that the conclusions remain the same as the original model that earlier brake, lower speed, and farther stopping distance, along with high clarity, visibility, safety and efficiency scores lead to safer work zone environment.

Table 5.1.2.4 Summary of Estimates of Simulator CFA Model

Latent Factor	Variable	Estimates		
		Lambda	Variance	Correlation
SimTrial_Safety	ZBrake	0.350	0.869	0.538
	ZSpeed	-0.353	0.867	
	ZStop	0.451	0.788	
SimSurvey_Safety	ZClarity	0.730	0.459	
	ZVisibility	0.882	0.213	
	ZSafety	0.947	0.094	
	ZEfficiency	0.697	0.506	

2. Modelling Field Study and Field Survey Data

When modeling the relationship between field survey data and field survey response, there were some difficulties. First, of the field MOEs, only vehicle speeds when passing 250 ft. distance marking (ft.) and full stop distance (ft.) were safety measures, the noncompliance data was binary and was not suitable for CFA. Second, the sample size of survey response was only 42, which was fairly small compared to the sample size of 334 of field data. Third, the survey design of field study was different from simulator survey, and asked the reason why survey respondents thought the devices were effective or not, instead of asking them to rate the four factors (clarity, visibility, safety, efficiency).

Although there were only two safety measures, they matched with the MOEs in simulator study, with first brake distance from work zone mission and the other two matched, which is considered as acceptable. CFA does not require variables in each factor match with each other.

As CFA requires sample size of every variable in the model need to be the same, or the missing data needs to be estimated, sedan data was extracted from field data. With missing data removed, sample size from sedan was 101. In order to math with the sample size of simulator study (115), another 14 samples were randomly selected from the pickup group. The reason why choosing sedan and pickup samples was because the AFAD simulator study was done using the car simulator, which was a compact sedan cab, and the driving behavior of similar sizes vehicle drivers could be similar.

For the features of clarity, visibility, safety and efficiency, values were assigned as rating score based on the score of effectiveness and reason for rating. When the device was rated as "very effective", and a factor was one of the reasons of rating, then this factor was assigned as 10. The corresponding scores for "effective", "neutral", "ineffective", "very ineffective" were eight, six, four and two. When a factor was not one of the rating reasons, then the score assignment would be nine, seven, five, three, and one. For example, if the MoDOT AFAD was rated as "very effective", with visibility, safety, efficiency chosen as reason for rating, then the scores of them were 10, 10, and 10. The remaining factor "clarity" had a score of nine. This assignment method may not be objective nor precise, however, it is the only way to match data with simulator data for synergy analysis.

Another issue about the field survey was, even with score assignment, the sample size was still lower than desired, as each survey response generated two groups of data because only MoDOT AFAD and human flagger were tested in the field, while four configurations were tested in simulator. The method used to make up was to randomly duplicate data points till sample size increased from 84 to 115.

The procedure of modeling was very similar to the model for simulator and post-simulator survey data. The path diagram is shown in Figure 5.1.2.2. As listed in Table 5.1.2.5, this model fits poorly. The estimates are reported in Table 5.1.2.6, but its reference value is low.

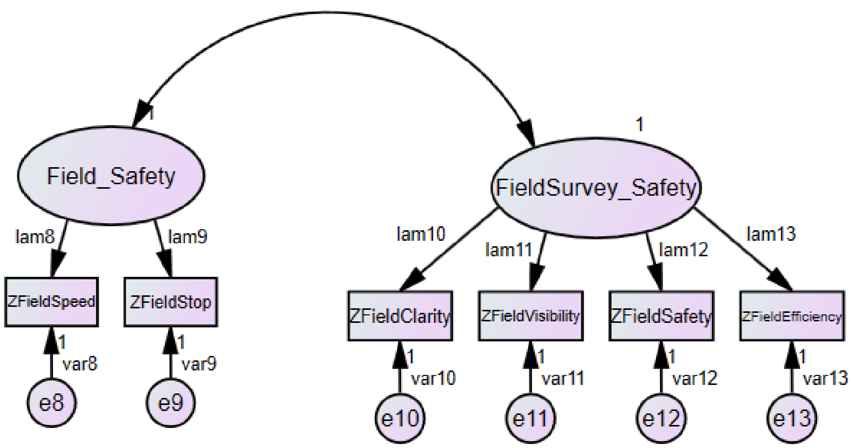


Figure 5.1.2.2 Path Diagram of Field Study and Field Survey Data

Table 5.1.2.5 Model Fit of Field Safety Model

Chi-Square	P-value	RMR	GFI	AGFI
99.562	0.000	0.352	0.839	0.577

Table 5.1.2.6 Summary of Estimates of Simulator CFA Model

Latent Factor	Variable	Estimates		
		Lambda	Variance	Correlation
Field_Safety	ZFieldSpeed	0.008	1.623	0.013
	ZFieldStop	4.153	-14.886	
FieldSurvey_Safety	ZFieldClarity	0.979	0.034	
	ZFieldVisibility	0.976	0.039	
	ZFieldSafety	0.978	0.035	
	ZFieldEfficiency	0.972	0.046	

With the current dataset, there is not much potential to improve the model for field data due to information missing and sample mix-match. The survey data and the field data did not have corresponding relationship, and a row of data did not come from the same person. If the data was going to be collected again, researchers need to have drivers stop after passing work zone and fill out the survey before they leave, which is not feasible and of low efficiency, which may cause terrible traffic congestion as well. Besides, time stamp and vehicle features need to be marked on the survey so that survey respondents could be identified from video recording and match with their vehicles, which would consume too much time.

3. Modelling Field and Simulator Driving Behavior Data

The path diagram of CFA model for field data and simulator data is shown in Figure 5.1.2.3. Data was standardized, as the other models, before feeding into estimates calculation. The model fit summary is provided in Table 5.1.2.7. The null hypothesis of consistency was rejected, which was not desirable. Although the AGFI was 0.71, which was not too low, the RMR was too high, and this model is not a very good fit. The estimates are reported in Table 5.1.2.8, and the results are consistent with researcher's assumption that earlier braking, lower speed, and farther full stop distance enhance work zone safety.

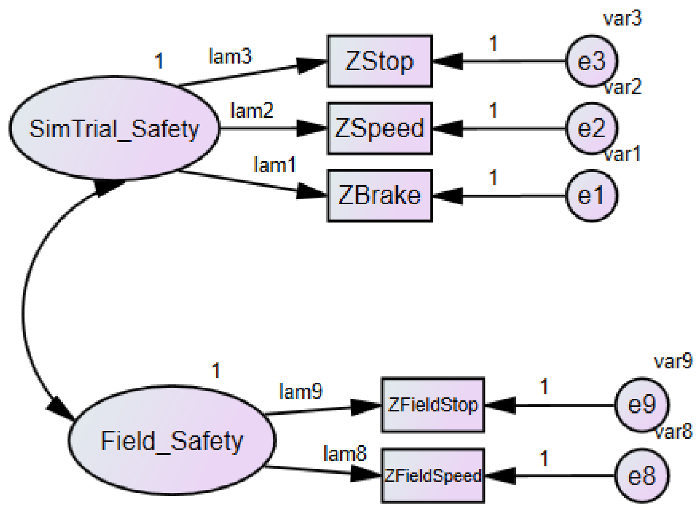


Figure 5.1.2.3 Path Diagram of Field and Simulator Driving Behavior Data

Table 5.1.2.7 Model Fit of Driving Behavior Safety

Chi-Square	P-value	RMR	GFI	AGFI
28.898	<0.001	0.208	0.923	0.710

Table 5.1.2.8 Summary of Estimates of Driving Behavior CFA Model

<table>
<tr><th rowspan="2">Latent Factor</th><th rowspan="2">Variable</th><th colspan="3">Estimates</th></tr>
<tr><th>Lambda</th><th>Variance</th><th>Correlation</th></tr>
<tr><td rowspan="3">SimTrial_Safety</td><td>ZBrake</td><td>0.275</td><td>0.920</td><td rowspan="5">-0.021</td></tr>
<tr><td>ZSpeed</td><td>-0.872</td><td>0.304</td></tr>
<tr><td>ZStop</td><td>0.229</td><td>0.943</td></tr>
<tr><td rowspan="2">Field_Safety</td><td>ZFieldSpeed</td><td>-5.702</td><td>-30.727</td></tr>
<tr><td>ZFieldStop</td><td>0.089</td><td>0.985</td></tr>
</table>

One way to improve the model is to ask the simulator participants to attend the field study, or invite the drivers who drove through the work zone to attend the simulator study, and record their driving behavior, then the corresponding relationship could be built and the field and simulator driving behaviors could be compared with each other. However, learning effect may exist and the latter driving behavior may be impacted.

4. Modelling Field and Post-Simulator Survey Responses

The model fit of field and post-simulator survey model is summarized in Table 5.1.2.9. The null hypothesis of consistency was failed to be rejected, which means the results are consistent with researcher's understanding or assumptions. The RMR is a little higher than 0.05, and it is sill acceptable as AGFI is high. Overall, this model is a good fit.

Table 5.1.2.9 Model Fit of Driving Behavior Safety

Chi-Square	P-value	RMR	GFI	AGFI
24.827	0.166	0.056	0.949	0.904

The estimates of field and post-simulator survey model are listed in Table 5.1.2.10, and visualized on path diagram in Figure 5.1.2.4. It is clearly indicated that the latent factor safety is able to be identified from the survey rating factors clarity, visibility, safety, and efficiency, in both simulator and field studies. In post-simulator survey, the direct rating of safety is the best indication of latent safety factor, and visibility follows it, then clarity. It is understandable that efficiency has a fairly lower coefficient estimate, as it is more of an operational feature. The estimates of field survey are very close, as they were calculated through score assignment and may be biased. One way to eliminate this bias is similar to the solution proposed in the driving behavior model. However, sequence bias or learning effect is still the main concern in this method.

Table 5.1.2.10 Summary of Estimates of Driving Behavior CFA Model

Latent Factor	Variable	Estimates		
		Lambda	Variance	Correlation
SimSurvey_Safety	ZClarity	0.729	0.460	-0.083
	ZVisibility	0.884	0.210	
	ZSafety	0.947	0.095	
	ZEfficiency	0.696	0.507	
FieldSurvey_Safety	ZFieldClarity	0.979	0.034	
	ZFieldVisibility	0.976	0.039	
	ZFieldSafety	0.978	0.035	
	ZFieldEfficiency	0.972	0.046	

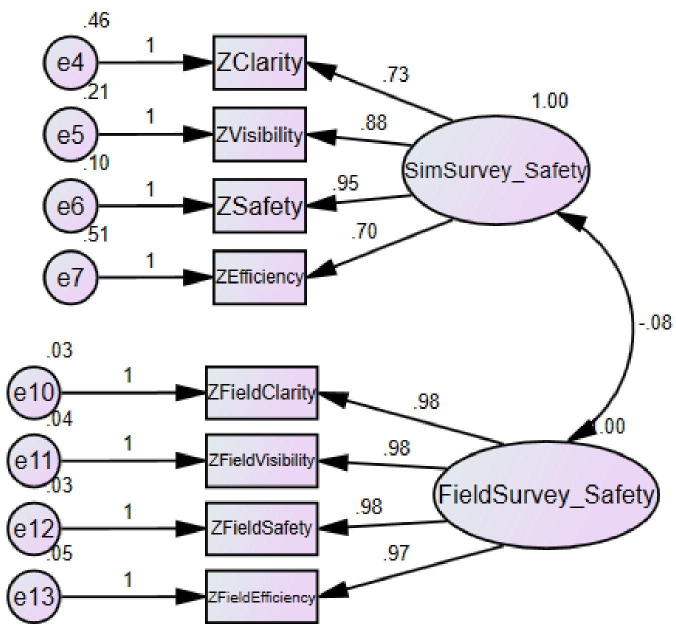


Figure 5.1.2.4 Path Diagram of Field and Simulator Survey Data

5. Modelling Field, Field Survey, Simulator and Post-Simulator Survey Data

A comprehensive CFA model is built to examine the synergy of field study, field survey, simulator study, and post-simulator survey. The MOEs of the studies were matched by two third, and the factors asked in questions were matched. The path diagram (Figure 5.1.2.5) of this model is more complicated than the other four aforementioned, as it is a four-factor CFA instead of two-factor, and correlations are considered exhaustively.

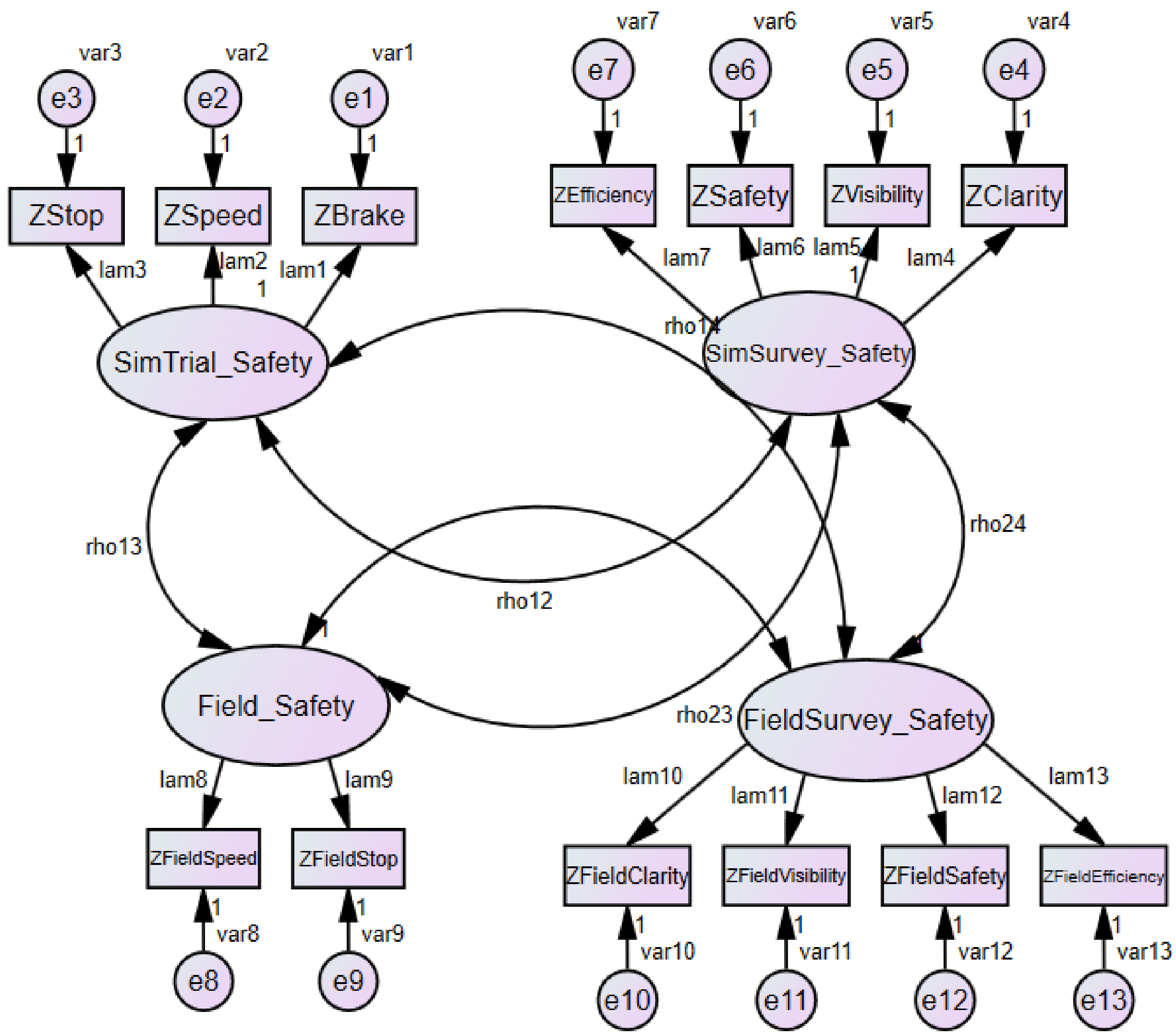


Figure 5.1.2.5 Synergy of Field Study, Field Survey, Simulator Study and Post-Simulator Survey

With prefix Z denotes data standardized, series of lam denotes coefficients lambdas, series of e denotes error term, and f1 being latent safety factor of simulator trials, f2 being latent safety factor of simulator survey, f3 being latent safety factor of field study, and f4 being latent safety factor of field survey, the formula set of this model is written as

ZBrake = lam1 f1 + e1
ZSpeed = lam2 f1 + e2
ZStop = lam3 f1 + e3
ZClarity = lam4 f2 + e4
ZVisibility = lam5 f2 + e5
ZSafety = lam6 f2 + e6
ZEfficiency = lam7 f2 + e7
ZFieldSpeed = lam8 f3 + e8
ZFieldStop = lam9 f3 + e9
ZFieldClarity = lam10 f4 + e10
ZFieldVisibility = lam11 f4 + e11
ZFieldSafety = lam12 f4 + e12
ZFieldEfficiency = lam13 f4 + e13

There are 32 parameters to be estimated, with 13 lambdas, 13 variance, and six correlations between each pair of latent factors. The model fit results are summarized in Table 5.1.2.11. The Chi-square is small, p-value is large, fail to reject null hypothesis, indicating that the results of model is consistent with researcher's prior assumption. RMR is within 0.05, and AGFI over 0.8, all these parameters are demonstrating a great model fit.

Table 5.1.2.11 Model Fit of Synergy

Chi-Square	P-value	RMR	GFI	AGFI
62.310	0.359	0.048	0.923	0.881

The summary of coefficient estimates is listed in Table 5.1.2.12, and the summary of correlations is listed in Table 5.1.2.13. From the estimate table, it could be interpreted that earlier brake, lower speed, farther full stop are factors enhancing work zone safety, in both field and simulator studies. Higher ratings of clarity, visibility, safety and efficiency are indicators for better work zone safety as well. Latent safety factors of simulator study and post-simulator survey have the highest correlation, which would be due to the corresponding relationship, as the data in one row is generated by the same person. However, none of the correlation is very high, and none of them could explain each other's variance well. To better illustrate the results, a result path diagram is shown in Figure 5.1.2.6.

Table 5.1.2.12 Summary of Estimates of Driving Behavior CFA Model

Latent Factor	Variable	Estimates	
		Lambda	Variance
SimTrial_Safety	ZBrake	0.383	0.845
	ZSpeed	-0.364	0.859
	ZStop	0.419	0.816
SimSurvey_Safety	ZClarity	0.728	0.461
	ZVisibility	0.881	0.215
	ZSafety	0.949	0.090
	ZEfficiency	0.695	0.508
Field_Safety	ZFieldSpeed	-0.723	0.468
	ZFieldStop	0.735	0.451
FieldSurvey_Safety	ZFieldClarity	0.979	0.034
	ZFieldVisibility	0.976	0.039
	ZFieldSafety	0.978	0.035
	ZFieldEfficiency	0.972	0.046

Table 5.1.2.13 Summary of Correlations

Variables		Correlation	Label
SimTrial_Safety	SimSurvey_Safety	0.530	rho12
SimTrial_Safety	Field_Safety	-0.140	rho13
SimTrial_Safety	FieldSurvey_Safety	-0.176	rho14
SimSurvey_Safety	Field_Safety	-0.215	rho23
SimSurvey_Safety	FieldSurvey_Safety	-0.085	rho24
Field_Safety	FieldSurvey_Safety	0.042	rho34

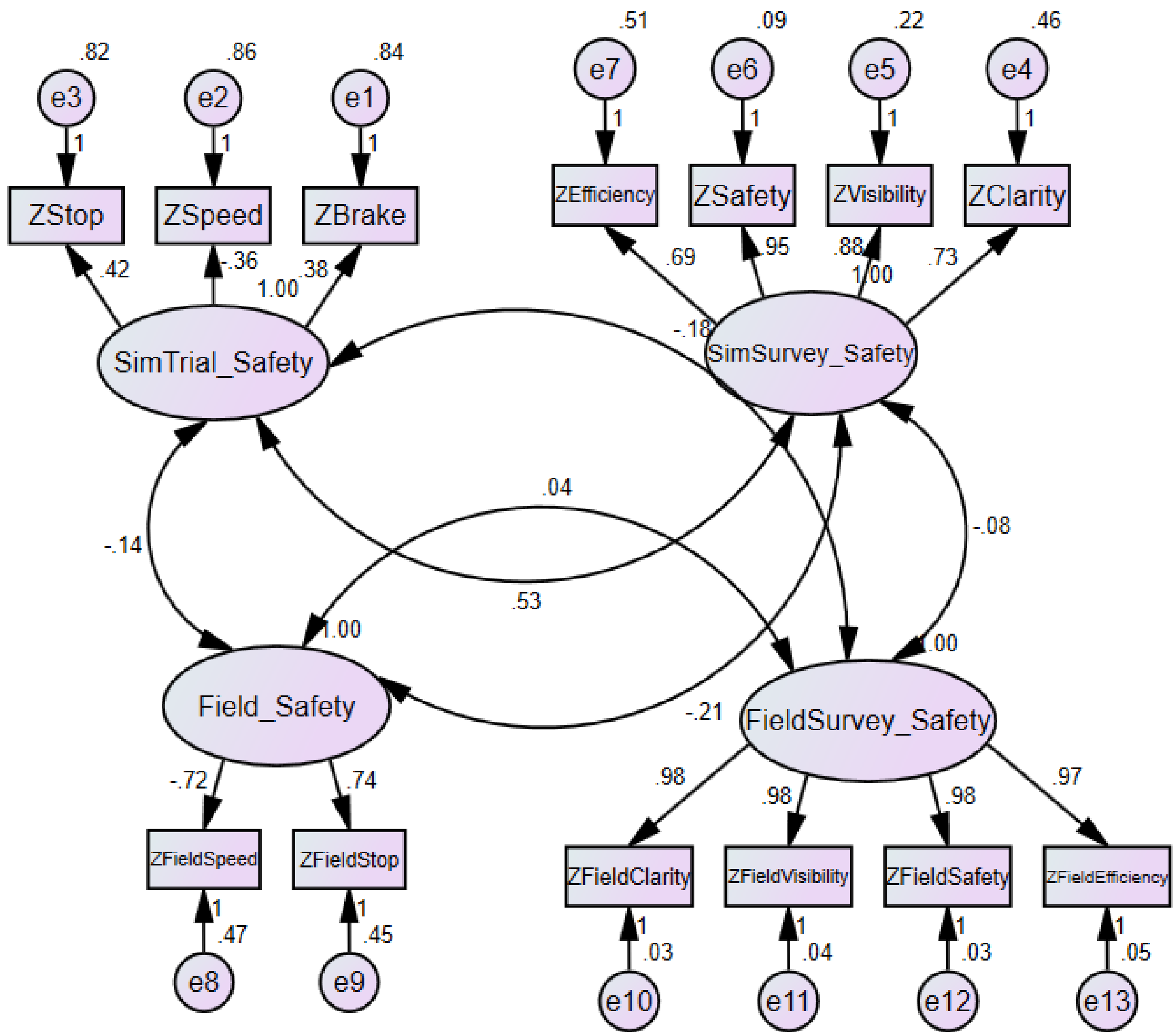


Figure 5.1.2.6 Estimates of Synergy of Field Study, Field Survey, Simulator Study and Post-Simulator Survey Model

5.2 MIXED MODELS TESTING FOR FACTORS IMPACTING DRIVING BEHAVIOR

Linear mixed models are an extension of simple linear model that allows both fixed and random effect, and are particularly used for non-independent and non-hierarchical data (Gałecki & Burzykowski 2013). Mixed models consider interactions between factors. In transportation engineering, mixed models could be applied when testing for multiple effects when interaction is unknown, or to find out the most impactful factor in people's driving behavior.

5.2.1 Modeling Driving Behavior and Demographic Information

Use the data collected from Green Light simulator studies as an example. In an effort to improve upon the traditional amber/white lights, the use of green lights on TMAs was investigated in a driving simulator. The study included the evaluation of four light color configurations: amber/white, green only, green/white, and green/amber. Each configuration was evaluated twice, once under daytime and once nighttime condition. Randomization of the scenario test order was used to limit possible sequence bias or order effect. The simulator trials included 29 participants. Merge distance (ft.) was captured to indicate safety feature, and farther distance is desired. Table 5.2.1.1 shows the properties information of data.

Table 5.2.1.1 Data information

Variable	Measurement Type	Purpose	Description
Distance	Continuous	Dependent variable	Merge distance captured
AgeGroup	Class	Factor	Two age groups (under and over 40)
ID	Class	Random	ID number of participants (29)
Gender	Class	Factor	Male and female
Configuration	Class	Factor, fixed	Major testing factor, 4 configurations
Time	Class	Factor, fixed	Day and night

In order to test for different factors influencing merge distance, multiple models were built to evaluate factor effects. Randomized block design is used to analyze the data. The easiest way to execute mixed model analysis it through SAS programing. The command “proc mixed” is the built-in function for mixed models. Four models are built in the example, and the SAS code is attached in Appendix L. In the models, Type III method was applied. A type III error is where the null hypothesis is rejected correctly but due to wrong reason. Tukey’s range test, also known as Tukey’s Honestly Significant Difference (HSD) (Gałecki & Burzykowski 2013), was also conducted to find the difference of the means of two groups in the same factor. The Akaike information criterion (AIC), which is out-of-sample prediction error (Akaike 1998), was adopted for model selection. A smaller AIC is more desired.

Model 1: Fixed effects of configuration and time

Model 1 tested only fixed effects of configuration and time, and the formula is presented as

$$Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$$

where α is configuration, i=1,2,3,4; β is time, j=1,2.

The result found that both configuration and time had significant effects on merge distance at 95% confidence level, as shown in Table 5.2.1.2. Based on Tukey results, vehicles merged significantly farther when encountered amber/white and green/amber than green/white and green only, and merge distance at nighttime was significantly farther away than daytime, as indicated in Table 5.2.1.3. It could be concluded that vehicles merged the earliest when they encountered amber/white TMA during nighttime. In this model, AIC = 3638.

Table 5.2.1.2 Type 3 Tests of Fixed Effects for Model 1

Effect	Num DF	Den DF	F Value	P-value
Configuration	3	227	6.27	<0.001*
Time	1	227	5.47	0.020**

* indicates statistical significance at 99% confidence level.
** indicates statistical significance at 95% confidence level.

Table 5.2.1.3 Differences of Least Squares Means of Model 1

Effect	Configuration	Time	Configuration	Time	Estimate	P-value
Configuration	Amber/white		Green Only		462.48	0.002*
Configuration	Amber/white		Green/amber		30.0345	0.996
Configuration	Amber/white		Green/white		339.22	0.045**
Configuration	Green Only		Green/amber		-432.45	0.005*
Configuration	Green Only		Green/white		-123.26	0.776
Configuration	Green/amber		Green/white		309.19	0.081
Time		Day		Night	-213.66	0.020**

* indicates statistical significance at 99% confidence level.
** indicates statistical significance at 95% confidence level.

Model 2: Fixed effects of configuration and time, and Random Effect of Participants

Model 2 considered not only the fixed effects of configurations and different time, but also the random effect caused by different experiment participants. The factor "ID" is added as a random effect in the model, as participants were selected randomly, and reaction of different people could be different due to physical and mental status. Formula is presented as

$$Y_{ijk} = \mu + \alpha_i + b_j + \beta_k + \varepsilon_{ijk}$$

where α is configuration, i=1,2,3,4; b is ID, j=1,2,…,29; β is time, k=1,2.

It turned out that the random effect is also significant, as shown in Table 5.2.1.4. The Tukey result is same as the previous model, as indicated in Table 5.2.1.5, amber/white and green/amber had significant farther merge distance than green/white and green only, and nighttime merge distance was farther than daytime. The conclusion remained the same: vehicles merged the earliest when encountered amber/white TMA during nighttime. AIC of Model 2 = 3556.7.

Table 5.2.1.4 Type 3 Analysis of Variance of Model 2

Source	DF	Expected Mean Square	Error Term	P-value
Configuration	3	Var(Residual) + Q(Configuration)	MS(Residual)	<.001*
Time	1	Var(Residual) + Q(Time)	MS(Residual)	0.002*
ID	28	Var(Residual) + 8 Var(ID)	MS(Residual)	<.001*
Residual	199	Var(Residual)	.	.

* indicates statistical significance at 99% confidence level.

Table 5.2.1.5 Differences of Least Squares Means of Model 2

Effect	Configuration	Time	Configuration	Time	Estimate	Pr > |t|	Adj P
Configuration	Amber/white		Green Only		462.48	<.001	<.001*
Configuration	Amber/white		Green/amber		30.0345	0.751	0.989
Configuration	Amber/white		Green/white		339.22	0.000	0.002*
Configuration	Green Only		Green/amber		-432.45	<.001	<.001*
Configuration	Green Only		Green/white		-123.26	0.194	0.562
Configuration	Green/amber		Green/white		309.19	0.001	0.007*
Time		Day		Night	-213.66	0.002	0.002*

* indicates statistical significance at 99% confidence level.

Model 3: Fixed effects of configuration, time, Age and Gender

As random effect of different participants is significant, there could be factors contributed to it. In this project, age and gender information was collected. The gender proportion was nearly balanced, with 57% male. However, the age distribution was skewed to the younger age. In order to gather more sample size for each age group, the original differentiation (16-25, 26-40, 41-55, 56-70) were merged into two groups: under 40, and above 41.

In Model 3, age and gender are added as fixed effects, and the formula is presented as

$$Y_{ijkm} = \mu + \alpha_i + \beta_j + \gamma_k + \tau_m + \varepsilon_{ijkm}$$

where α is configuration, i=1,2,3,4; β is time, j=1,2; γ is gender, k=1,2; τ is age group, m=1,2. However, result showed that neither age nor gender have significant impact on the merge distance, as shown in Table 5.2.1.6.

Table 5.2.1.6 Type 3 Tests of Fixed Effects for Model 3

Effect	Num DF	Den DF	F Value	P-value
Configuration	3	225	6.26	<0.001*
Time	1	225	5.47	0.020*
Gender	1	225	0.04	0.842
AgeGroup	1	225	1.82	0.1786

* indicates statistical significance at 99% confidence level.

Model 4: Fixed effects of configuration, time, Age and Gender with All Possible Interactions

There is possible interaction between the factors. Although each configuration was tested once in daytime and once in nighttime and may not be considered as replicates, repeated measure was performed. All possible interactions are added based on Model 3. The formula is

$$Y_{ijkm} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \gamma_k + \alpha\gamma_{ik} + \beta\gamma_{jk} + \alpha\beta\gamma_{ijk} + \tau_m + \alpha\tau_{im} + \beta\tau_{jm} + \gamma\tau_{km} + \alpha\beta\tau_{ijm} + \alpha\gamma\tau_{ikm} + \beta\gamma\tau_{jmk} + \alpha\beta\gamma\tau_{ijkm} + \varepsilon_{ijkm}$$

where α is configuration, i=1,2,3,4; β is time, j=1,2; γ is gender, k=1,2; τ is age group, m=1,2.

The results in Table 5.2.1.7 showed that the only significant interaction was between gender and age. Other than this, there was no significant interactions. When interactions were considered, the main effects of configurations, gender, and age were significant, however, the main effect of time was no longer significant. AIC = 3231.7 for this model.

Table 5.2.1.7 Type 3 Tests of Fixed Effects for Model 4

Effect	P-value
Configuration	0.008*
Time	0.086***
Configuration*Time	0.151
Gender	0.002*
Configuration*Gender	0.404
Time*Gender	0.302
Configur*Time*Gender	0.316
AgeGroup	0.002*
Configurati*AgeGroup	0.840
Time*AgeGroup	0.630
Configu*Time*AgeGrou	0.194
Gender*AgeGroup	<.001*
Config*Gender*AgeGroup	0.592
Time*Gender*AgeGroup	0.901
Conf*Time*Gend*AgeGroup	0.597

* indicates statistical significance at 99% confidence level.
*** indicates statistical significance at 90% confidence level.

The Tukey result in Table 5.2.1.8 showed that when potential interactions were considered, amber/white had significant farther merge distance than green/white and green only, but the difference between amber/white and green/amber could not be told; difference between daytime and nighttime was not significant; male had farther merge distance than female, and younger aged drivers (under 40) reacted earlier than the older drivers (over 40).

Table 5.2.1.8 Differences of Least Squares Means of Model 4

Effect	Configuration	Time	Gender	AgeGroup	Configuration	Time	Gender	AgeGroup	Estimate	Adj P
Configuration	Amber/white				Green Only				640.19	0.011**
Configuration	Amber/white				Green/amber				126.20	0.926
Configuration	Amber/white				Green/white				434.23	0.149
Configuration	Green Only				Green/amber				-513.99	0.061***
Configuration	Green Only				Green/white				-205.96	0.745
Configuration	Green/amber				Green/white				308.03	0.436
Time		Day				Night			-249.45	0.086***
Gender			F				M		-460.82	0.002*
AgeGroup				Over 40				Under 40	462.93	0.002*

* indicates statistical significance at 99% confidence level.
** indicates statistical significance at 95% confidence level.
*** indicates statistical significance at 90% confidence level.

Summary of Four Mixed Models

To summarize, drivers reacted to amber/white TMA and merged the earliest, which is the safest configuration when only merge distance was considered and measured. Although Model 4 had the smallest AIC, it is the most complicated, and may not be legitimate as the sample size was small. Model 2 had the second smallest AIC, and it took random effect into consideration, and it should be regarded as the most appropriate model for this experiment. So, configuration and time fixed effects and random effect of participants should be considered in the model, as all of them are significant.

The biggest short coming of this dataset is small sample size and no replicates. The small sample size what due to time constrain, as researchers would like to get simulator results quicker, in order to determine the configurations to be tested in the field, and the sponsors would need the overall results and deploy the device on roadways when construction season begins. Each configuration was tested once in daytime and once in nighttime, without replicates, which was because previous research showed that a simulator test within 20 minutes generated less discomfort.

If unlimited resource is provided, more participants will be recruited, at least 30 for each age group; gender will be balanced; for each configuration, it should be tested three times in daytime and three times in nighttime to provide replicates in order to test for interactions.

5.2.2 Modeling Driving Behavior and Driver's Perspective and Understanding of New Designs

Mixed models are applicable to help understand the relationship between driver's perspective and understanding and the driving behavior as well. Use the simulator data and post-simulator survey response as an example. A model was built to test for the relationship among driver's understanding of traffic control devices (MoDOT AFAD denoted as AFAD1, AFAD with alternative sign denoted as AFAD2, AFAD without CMS denoted as AFAD3, and human flagger denoted as HF), rating of features in configurations and simulator fidelity, and the full stop distance from work zone. Table 5.2.6.1 indicates that the type of devices and driver's understanding of meaning of devices had significant impacts on how faraway they came to a full stop before the work zone, and both confidence level were 95%. The rating of efficiency and highway fidelity were significant at 90% confidence level.

Table 5.2.2.1 Type 3 Tests of Fixed Effects for Modeling Full Stop Distance (ft.) of AFAD Simulator Study

Source	DF	Expected Mean Square	Error Term	Error DF	P-value
Device	3	Var(Residual) + Q(Device)	MS(Residual)	98	0.038**
Age	4	Var(Residual) + Q(Age)	MS(Residual)	98	0.824
Gender	1	Var(Residual) + Q(Gender)	MS(Residual)	98	0.596
Understanding	1	Var(Residual) + Q(Understanding)	MS(Residual)	98	0.033**
Clarity	1	Var(Residual) + Q(Clarity)	MS(Residual)	98	0.587
Visibility	1	Var(Residual) + Q(Visibility)	MS(Residual)	98	0.846
Safety	1	Var(Residual) + Q(Safety)	MS(Residual)	98	0.381
Efficiency	1	Var(Residual) + Q(Efficiency)	MS(Residual)	98	0.095***
CMSNecessary	1	Var(Residual) + Q(CMSNecessary)	MS(Residual)	98	0.696
OnHighway	1	Var(Residual) + Q(OnHighway)	MS(Residual)	98	0.061***
DriveFreely	1	Var(Residual) + Q(DriveFreely)	MS(Residual)	98	0.109
Residual	98	Var(Residual)	.	.	.

** indicates statistical significance at 95% confidence level.
*** indicates statistical significance at 90% confidence level.

Table 5.2.2.2 Differences of Least Squares Means of Model 4

Effect	Device	Understanding	Device	Understanding	P-value
Understanding		0		1	0.033**
Device	AFAD1		AFAD2		0.688
Device	AFAD1		AFAD3		0.097***
Device	AFAD1		HF		0.007*
Device	AFAD2		AFAD3		0.159
Device	AFAD2		HF		0.008*
Device	AFAD3		HF		0.087***

* indicates statistical significance at 99% confidence level.
** indicates statistical significance at 95% confidence level.
*** indicates statistical significance at 90% confidence level.

It is implied that the configuration of a traffic control device and how drivers understand the meaning of it made significant difference on driving behavior. Public agencies need to ensure that device designs and message displayed are effectively conveying easily understandable information, and enhance public education before deploying any new designs.

5.2.3 Other Applications of Mixed Models in Transportation Engineering

The mixed model verifies significant factors and identifies correlation between variables, pointing out possible weights of factors in driving behavior. This method is applicable for determining significant features in new road design, new device configurations, or identifying the most important considerations in tradeoff situations.

Results of mixed model could be treated as a pre-selection procedure of regression modeling as well. When there are too many variables in a regression model, dropping off the less representative ones would improve calculation efficiency, make the model more interpretable.

5.3 MODELING SIMULATOR LEARNING EFFECTS

5.3.1 Paired T-test of Repeated Measures

A paired t-test is one-sample t-test, with the difference of each pair constructed as the group, and the mean of the group is the mean of the difference. Paired t-test in transportation engineering could be utilized when education occurs, for example, participants drive the car simulator with new signage design, then take a break and be educated about the new signage, then drive again under the same scenarios. Researchers collect data before and after education and compare them, to see if the effect of education is significant. Or paired t-test could be used to identify learning effect and sequence bias. Participants take two trials of the same designs and compare the results, and see if there is different driving behavior between first and second trials.

The null hypothesis of one sample t-test is the mean of this group is not significantly different than a numeri number μ_0, or the difference between the mean of this group and the specific numerical number μ_0 is not significantly different than zero. The formula is presented as:

$$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$$

where $\bar{X}$ is the mean of the group, s is the standard deviation of the group, and n is the sample size of the group.

Table 5.3.1.1 is an example of data comparison for two simulator trials in AFAD simulator study. In each trial, participants encountered three AFAD configurations and one human flagger in randomized order, and first brake distance, speed at 250 ft. from work zone, and full stop distance were recorded. The paired t-test results indicated that participants hit the brake later in second trial than first trial, however, the difference was not significant. The speed of car when it passed 250 ft. marking cone for second trial was significantly higher than the first time. Participants came to full stop later in the second trial than the first one, and the difference was significant at 90% confidence level.

Table 5.3.1.1 Example of AFAD Simulator Studies Data Analysis Using Paired T-Test

	First Brake Distance from Work Zone (ft.)		Speed at 250 ft. from Work Zone (mph)		Full Stop Distance from Work Zone (ft.)	
	Trial 1	Trial 2	Trial 1	Trial 2	Trial 1	Trial 2
Mean	321.56	310.30	25.94	30.63	82.82	75.70
Standard Deviation	119.74	107.95	11.97	12.31	44.88	45.85
Diff (Trial 2-Trial1)	-11.26		4.69		-7.13	
P-value of t-test	0.213		<0.001*		0.097***	

* indicates statistical significance at 99% confidence level.
*** indicates statistical significance at 90% confidence level.

From the results above, it could be implied that there was a learning effect in simulator experiments. As participants keep driving, they were more familiar with the car simulator, and tended to be more relaxed in the second trial. So, they were not as careful as the first trial, and approached the work zone less defensively.

5.3.2 Discrepancy between Driver's Understanding and Driving Behavior

Empirically, the driving behavior may vary from driver's understanding. For example, drivers may run the red signal when driving impaired or distractively. When asked about their understanding of a new type of traffic signal or signage design, drivers may be able to provide a correct answer, but when they encounter the same design, violation or noncompliance may happen.

In AFAD simulator study, participants were asked to drive to a clinic. On their way, they would drive through eight rural work zones, with each work zone having one type of traffic control device. On their way, they would encounter four types at randomized orders, and for each type, they would encounter twice. There was no break or education between two trials. In the post-simulator survey, participants were asked for four multiple choice questions, with each question test for correct understanding of one device.

There were 32 participants, and four answers from each, then the total collected answers were 128, with 112 correct answers and 16 incorrect answers. Table 5.3.2.1 shows the result of the inconsistency between driver's answer of understanding and their driving behavior. Within the 112 answers of understanding the meaning of traffic control devices correctly, correct driving behavior of first trial was 103 (80.5%), and the for the second trial was 110 (85.9%), which means seven participants corrected their decision making when they encountered the work zone the second time. Within the 16 incorrect answers, there was only one noncompliance driving behavior at the first trial, and in the second trial, there was no noncompliance.

Table 5.3.2.1 Inconsistency of Driver's Understanding and Driving Behavior

Understand	Non-compliance	Consistent	Count		Percentage (%)		
			Trial 1	Trial2	Trial 1	Trial2	Difference
Yes	No	Yes	103	110	80.47	85.94	5.47
Yes	Yes	No	9	2	7.03	1.56	-5.47
No	No	No	15	16	11.72	12.50	0.78
No	Yes	Yes	1	0	0.78	0	-0.78
Total			128	128	100	100	N/A

The reason why the noncompliance rate decreased at the second trial would be a good indicator of simulator learning effect. Participants were more familiar with the driving simulator environment, physically and virtually, then they gained the knowledge from experience, and obeyed the instruction. Or the first time of noncompliance was due to novelty effect, which was not effective anymore at the second trial.

Regarding the discrepancy between understanding and driving behavior, there are some interpretation and assumptions. For those who answered the questions correctly but did not follow the instruction, it may be because the post-simulator survey was taken after the trials were completed, and due to learning effect, participants were able to understand the meaning of the devices after driving through it twice. For those who answered the question incorrectly but obeyed the traffic control, it may be due to discrepancy between mind and actual behavior, which is under the category of behavioral psychology.

It could be implied that sometimes survey response alone is not a legitimate source of information, as mind behavior discrepancy occurs. In transportation engineering, conducting simulator or field studies would help to acquire quantitative data. With the combination of scientific studies and surveys, both objective and subjective, and quantitative and qualitative results could be obtained, and conclusions would be more reliable and convincing.

CHAPTER 6: CONCLUSIONS AND DISCUSSIONS

Original and significant contribution of this dissertation focuses on investigation of smart work zone technologies using mixed simulator and field studies to improve safety & efficiency, of which methodology includes simulator studies design, survey design, synergy of simulator and field studies, utilization of psychophysics and biofeedback utilization in simulators, and application of statistical modeling in transportation engineering incorporating simulator and field acquired quantitative and qualitative performance measures.

The synergy of simulator and field studies is a key. Simulator provides a safe and cost-effective environment to test new designs, and scenarios are able to be coded at will, which is sometimes impossible in the real world. The advantage of field study over simulator study is that field study obtains realistic data and captures absolute validity, while simulator study acquires relative validity. Mixed simulator and field studies work well together, as field study either calibrates simulator design at prior, or validate simulator data afterwards.

A new configuration of Automated Flagger Assistance Devices (AFADs), a traffic control device for stationary work zone replacing human flagger, was proposed by MoDOT, innovatively combining stop/Slow paddle with red/yellow beacon, adding a changeable message sign attached to a Truck-Mounted-Attenuator. This study contained a field study followed by survey, and a simulator study followed by post-simulator survey. It is the first scientific study of AFAD collecting quantitative data in the United States, and the first simulator study of AFADs in the world. Field study was conducted before simulator study to calibrate simulator study design, and both results were consistent with each other that AFAD had the potential to improve work zone safety by protecting flaggers, slowing down traffic, and lengthening full stop distance. Both simulator and field surveys indicated that drivers/participants preferred AFADs more due to its high visibility and clear guidance. MoDOT decided to start implementing AFADs to statewide rural work zones in Fall 2017. AFAD is transferable to other states due to its safety features and relative validity compared with human flagger, mobility, and effectiveness.

Factor analysis of AFAD project data confirmed the assumption of latent safety factor. Earlier braking, lower approaching speed, and stopping farther away from work zone are confirmed to enhance work zone safety, which is consistent with intuition and prior knowledge. Mixed model analysis indicated the significant effect of driver's understanding of signage or traffic control devices on driving behavior. However, discrepancy between driver's understanding and driving behavior does exist. Repeated measure of simulator trials implies a learning effect, and better performance was observed from latter trials.

Different light color combinations containing green lights on Truck-Mounted Attenuators (TMAs) were tested to enhance visibility and catch driver' attentions, aiming to improve mobile work zone safety and reduce TMA hits. A simulator study followed by a post-simulator survey tested four configurations of light bars on TMAs: amber/white, green only, green/amber and green/white. A field study was conducted after simulator study to validate results and obtain absolute validity. MoDOT was the first state proposing using green as work zone alert color, and this project was the first quantitative study on green light TMAs. An inverse relationship between visibility/awareness of work zone and arrow board recognition/easy on eyes was derived from the simulator results based on eye tracking data, and green/amber combination was ranked the highest as a compromise by participants. However, green/amber was not test in field study due to lack of equipment. Factors found in field study leading to lower bypassing vehicle speeds were slow TMAs and nighttime. Extend mixed model analysis of simulator study found an interaction of gender and age that impacted driving behavior. Nighttime scenario captured significant slower vehicle bypassing speed in simulator study, which is consistent with field study. Overall, simulator and field results did not point in a single direction, and all light color combination configurations appeared to be viable. MoDOT decided to stop further investigation on green light TMAs, and remain using amber/white TMAs for mobile work zones. If study is repeated at a different state or a different country, the result may be different due to different testing season and natural environment, traffic volume difference, and potential aggressive driving behavior.

The combination of Exploratory Factor Analysis and Confirmatory Factor Analysis is appropriate in transportation data analysis, as it is able to uncover the underlaying relationship among variables, identify safety features, and confirm prior assumptions and theories. Besides safety, factor analysis is also suitable for operational data modeling such as identifying efficiency features, transportation project and risk management, and public surveys.

Mixed model fits mixed simulator and survey result analysis, as it is able to figure out the factors that significantly impact driving performance, and help to interpret the relationship between drivers' demographic information and driving behaviors. It is also suitable for finding out relationship between driver's understanding and perspective about designs and their driving performance.

In conclusion, investigating smart work zone technologies using mixed simulator and field studies is a new preferable method, as it balances safety, cost-effectiveness, and veracity. With the auxiliary of psychophysical devices and statistical analysis, the synergy of simulator and field studies conducts research scientifically, obtains quantitative and qualitative data, which could be analyzed through factor analysis, mixed models and other statistical methods, and delivers results legitimately.

REFERENCES

AASHTO (2009). "Guidelines for the Selection and Application of Warning Lights on Roadway Operations Equipment."Washington, D.C.

AASHTO (2013). "A Policy on Geometric Design of Highways and Streets."

Ackerley, N., Sertkaya, A., & Lange, R. (2010). “Food transportation safety: characterizing risks and controls by use of expert.” *Food protection trends, 30*(4), 212-222.

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. *Selected papers of hirotugu akaike* (pp. 199-213). Springer, New York, NY.

Alack v. Vic Tanny International of Missouri. (1996). No. 78423, 923 S.W.2d 330, Supreme Court of Missouri, En Banc.

Antonucci, N., Hardy, K., Bryden, J., Neuman, T., Pfefer, R., and Slack, K. (2005). "NCHRP Report 500: Guidance for Implementation of the AASHTO Strategic Highway Safety Plan—Volume 17: A Guide for Reducing Work Zone Collisions." *Transportation Research Board of the National Academies,* Washington, DC.

ATSSA (2012). "Guidance on the Use of Automated Flagger Assistance Devices."

Brown, H., Sun, C., and Cope, T. (2015). "Evaluation of Mobile Work Zone Alarm Systems." *Transportation Research Record: Journal of the Transportation Research Board* (2485), 42-50.

Brown, M., Avery, P., and Sturgeon, P. (2014). “Enhanced Work Zone Safety Through Cooperative Autonomous Vehicle Systems.” *2014 International Conference on Connected Vehicles and Expo (ICCVE).* IEEE, 2014.

Brown, T. A. (2014). *Confirmatory factor analysis for applied research*. Guilford Publications.

Bundele, M. M., and Banerjee, R. (2009). “Detection of fatigue of vehicular driver using skin conductance and oximetry pulse: a neural network approach.” In *Proceedings of the 11th International Conference on Information Integration and web-based applications & services* (pp. 739-744). ACM.

Cohen, J. (1977). "Statistical power analysis for the behavioural sciences (Revised edition)." *New York*, 7.

Cook, S., Quigley, C., and Clift, L. (2000). "Motor vehicle and pedal cycle conspicuity: part 3-vehicle mounted warning beacons. Summary report."

Cottrell Jr, B. (2006). "Evaluation of the Autoflagger in Virginia."

Cottrell Jr, B. H. (2015). "Investigation of Truck Mounted Attenuator (TMA) Crashes in Work Zones in Virginia."

De Winter, J. C., and Dodou, D. (2010). "Five-point Likert items: t test versus Mann-Whitney-Wilcoxon." *Practical Assessment, Research & Evaluation*, 15(11), 1-12.

FHWA (2005a). "Technical Provisions for AFADs." U. S. Department of Transportation.

FHWA (2005b). "Revised Interim Approval for the use of Automated Flagger Assistance Devices in Temporary Traffic Control Zones", U. S. Department of Transportation.

FHWA (2009). "Manual on Uniform Traffic Control Devices for Streets and Highways." F. H. Administration, ed. Washington, D.C.

FHWA (2012). “Non-Motorized User Safety.” U. S. Department of Transportation.

FHWA (2016). “Design and Operation of Work Zone Stategies to Improve Large Truck Safety.” U. S. Department of Transportation.

FMCSA (2014). “Commercial Motor Vehicle Traffic Safety Facts.” U. S. Department of Transportation.

Finley, M. (2013). "Field Evaluation of Automated Flagger Assistance Devices in Work Zones on Two-Lane Roads." *Transportation Research Record: Journal of the Transportation Research Board* (2337), 1-8.

Finley, M. D., Ullman, B. R., Trout, N. D., and Park, E. S. (2011). "Studies to determine the effectiveness of automated flagger assistance devices and school crossing devices." Texas Transportation Institute.

Ford, J. K., MacCallum, R. C., & Tait, M. (1986). “The application of exploratory factor analysis in applied psychology: A critical review and analysis.” *Personnel Psychology, 39,* 291–314.

Frank, D. L., Khorshid, L., Kiffer, J. F., Moravec, C. S., and McKee, M. G. (2010). “Biofeedback in medicine: who, when, why and how?.” *Mental health in family medicine, 7(2),* 85–91.

Frey, J. H. (1989). *Survey research by telephone* (2nd ed., pp. 33-78). Newbury Park, CA: Sage.

Friedman, M. (1940). A comparison of alternative tests of significance for the problem of m rankings. *The Annals of Mathematical Statistics, 11(1),* 86-92.

Gałecki, A., & Burzykowski, T. (2013). “Linear mixed-effects model.” *Linear Mixed-Effects Models Using R* (pp. 245-273). Springer, New York, NY.

Garrison, W. L., & Marble, D. F. (1964). “Factor-analytic study of the connectivity of a transportation network.” *Regional Science Association* 12(1), 231-238.

Gaudeau, C., Golding, J. F., Thevot, F., Lucas, Y., Bobola, P., and Thouvenot, J. (2005). "Space Motion Sickness and Stress Training Simulator using Electrophysiological Biofeedback." *Topical Teams in Life & Physical Sciences: Towards New Research Applications* (Vol. 1281, pp. 136-147).

Gibbons, R. B. (2008). Selection and Application of Warning Lights on Roadway Operations Equipment, Transportation Research Board.

Girzadas Jr, D. V., Delis, S., Bose, S., Hall, J., Rzechula, K., & Kulstad, E. B. (2009). "Measures of stress and learning seem to be equally affected among all roles in a simulation scenario." Simulation in Healthcare, 4(3), 149-154.

Google Maps. (2017). https://maps.google.com/.

Hager, M., Wilson, S., Pollak, T., and Rooney, P. (2003). "Response Rates for Mail Surveys of Nonprofit Organizations: A Review and Empirical Test." Nonprofit and Voluntary Sector Quarterly 32(2), 252-267.

Harman, H. H. (1976). "Modern factor analysis." University of Chicago press.

Harris, L. (2002). "RC Flagman Gets Test Run in Washington County." KUTC Newsletter, University of Kansas, Kansas.

Hogg, R. V., Tanis, E. A., & Zimmerman, D. L. (1977). Probability and statistical inference (Vol. 993). New York: Macmillan.

Howell, B. K., Pigman, J. G., and Agent, K. R. (2015). "Work Vehicle Warning Lights: Color Options and Effectiveness."

IntelliStrobe (2017). "IntelliStrobe Safety Systems." (January 05, 2017).

IRB (Institutional Review Board). (2015). "Informed Consent – Types and Elements." IRB 46.116, University of Missouri-Columbia, MO.

Ishak, S., Qi, Y., and Rayaprolu, P. (2012). "Safety evaluation of joint and conventional lane merge configurations for freeway work zones." *Traffic injury prevention*, 13(2), 199-208.

Jessberger, R. (1999). "RC Flagman Evaluation." Ohio Department of Transportation.

Jolliffe, I. T., and Cadima, J. (2016). "Principal component analysis: a review and recent developments." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374*(2065), 20150202.

Kaiser, H. F. (1958) The varimax criterion for analytic rotation in factor analysis. *Psychometrika, 23*(3), 187-200.

Kaptein, N., Theeuwes, J., and Van Der Horst, R. (1996). "Driving simulator validity: Some considerations." *Transportation Research Record: Journal of the Transportation Research Board* (1550), 30-36.

Kennedy, R. S., Lane, N. E., Berbaum, K. S., and Lilienthal, M. G. (1993). "Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness." *The international journal of aviation psychology*, 3(3), 203-220.

Kline, R. B. (2015). *Principles and practice of structural equation modeling*. Guilford publications.

Knefel, E. (1994). “StarCraft lawsuit: Privacy and piracy.” CNN., <http://www.cnn.com/TECH/computing/9805/04/privacy_piracy/> (Dec. 2, 2015).

Kurniawan, H., Maslov, A. V., and Pechenizkiy, M. (2013). “Stress detection from speech and galvanic skin response signals.” In *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems* (pp. 209-214). IEEE.

Lee, J., Seo, Y., Zhang, W., and Wettergree, D. (2013). “Kernel-Based Traffic Sign Tracking to Improve Highway Workzone Recognition for Reliable Autonomous Driving.” *Proc, the 16th International IEEE Annual Conference on Intelligent Transportation Systems,* Hague, Netherlands.

MacLean, D., Roseway, A., and Czerwinski, M. (2013, May). MoodWings: a wearable biofeedback device for real-time stress intervention. In *Proceedings of the 6th international conference on PErvasive Technologies Related to Assistive Environments* (p. 66). ACM.

Milton, J., & Arnold, J. (1995). “Introduction to probability and statistics: principles and applications for engineering and the computer sciences.”

Minnesota Department of Transportation (MnDOT). (2005). "AUTOFLAGGER Research Project Final Report."

Missouri Department of Transportation (MoDOT). (2018). "Centerline and/or Laneline Striping on Multi-lane, Undivided Highways - MT." *616.8.35a (TA-35a)*, MoDOT, ed.

Missouri Department of Transportation (MoDOT). (2006). "District 9 Pilots New Flagging System."

Missouri Department of Transportation (MoDOT). (2009). "Engineering Policy Guide." http://epg.modot.org/index.php?title=Main_Page. (July 1, 2017).

Missouri Department of Transportation (MoDOT). (2017). “Innovation Showcase. Ticket Cost Changeable Message Sign.” https://www.modot.org/sites/default/files/documents/TicketCostCMS.pdf (May 04, 2020).

Missouri Department of Transportation (MoDOT). (2019). “Provide Outstanding Customer Service Tracker.”

Missouri Department of Transportation (MoDOT). (2008). “Staff Summary. Evaluation of Automatic Flagger Assistance Devices.” https://spexternal.modot.mo.gov/sites/cm/CORDT/ss08005.pdf (May 04, 2020).

Nguyen, A. T., Nguyen, L. D., Le-Hoai, L., & Dang, C. N. (2015). “Quantifying the complexity of transportation projects using the fuzzy analytic hierarchy process.” *International journal of project management, 33*(6), 1364-1376.

NHTSA (2019). “Distracted Driving.” <https://www.nhtsa.gov/risky-driving/distracted-driving>

North America Traffic (2016). "RCF 2.4 Automated Flagger Assistance Device." <http://northamericatraffic.com/flagging-devices/rcf-2-4/>. (January 05, 2017).

Lahr v. National Transp. Safety Bd. (2006). No. 453 F. Supp.2d 1153, United States District Court, C.D. California.

Likert, R. (1932). "A Technique for the Measurement of Attitudes". *Archives of Psychology*. 140: 1–55.

ODOT (2013). "ODOT's New Colored Light Combinations on Snow Removal Equipment." <http://www.dot.state.oh.us/Services/Pages/New-Colored-Light-Combinations-on-Snow-Removal-Equipment.aspx>.

Perreault, W. D. (1975). "Controlling order-effect bias." *The Public Opinion Quarterly*, 39(4), 544-551.

Pigman, J. G., Agent, K. R., and Green, E. R. (2006). "Evaluation of work zone safety operations and issues."

Preedy, V. R. (2010). *Handbook of disease burdens and quality of life measures* (pp. 4271-4272). R. R. Watson (Ed.). New York: Springer.

Pop-Jordanova, N., and Gucev, Z. (2010) “Game-based peripheral biofeedback for stress assessment in children.” *Pediatrics International*, Vol. 52, No. 3, pp. 428-431.

Popuri, Y., Proussaloglou, K., Ayvalik, C., Koppelman, F., & Lee, A. (2011). “Importance of traveler attitudes in the choice of public transportation to work: findings from the Regional Transportation Authority Attitudinal Survey.” *Transportation, 38*(4), 643-661.

Richards, S. H., and Bowman, J. (1981). "Driver understanding of work zone flagger signals and signaling devices."

Safety Technologies (2015a). "AUTOFLAGGER 76." <http://www.autoflagger.com/products/autoflagger-76/>. (December 28, 2016).

Safety Technologies (2015b). "AUTOFLAGGER 54." <http://www.autoflagger.com/products/autoflagger-54/>. (December 28, 2016).

Schoelz, M., Sun, C., Qing, Z., and Edara, P. (2019). “Federated Simulator Study of External Communication Displays for Connected Truck Platoons.” *Proc, 2020 Transportation Research Board Annual Meeting*, Washington D.C.

Srinivasan, S., Carrick, G., Heaslip, K., and Washburn, S. (2007). "Analysis of crashes in freeway work-zone queues: a case study." *Research Project*.

Subramaniam, S. K., Ganapathy, V. R., Subramonian, S., and Hamidon, A. H. (2010). "Automated traffic light system for road user’s safety in two lane road construction sites." *WSEAS Transactions on Circuits and Systems*, 2(9), 71-80.

Syed, S. I., & Khan, A. M. (2000). “Factor analysis for the study of determinants of public transit ridership.” *Journal of Public Transportation, 3*(3), 1.

Synergy Technology (2017). "Automated Flagger AF-100." <http://www.noflaggers.com/automated_flagger_overview.html>. (January 05, 2017).

Terhaar, E. F. (2014). "Implementation of Automatic Flagger Assistance Devices (AFADs) for Minnesota Department of Transportation Flagger Operations." Minnesota Department of Transportation, Research Services & Library.

Trout, N., Finley, M., and Ullman, B. (2013). "Motorists' Understanding of Automated Flagger Assistance Devices in Work Zones." *Transportation Research Record: Journal of the Transportation Research Board* (2337), 42-49.

Trout, N. D., and Ullman, G. L. (1997). "Devices and Technology to Improve Flagger/Worker Safety."

Villarejo, M. V., Zapirain, B. G., and Zorrilla, A. M. (2012). “A stress sensor based on Galvanic Skin Response (GSR) controlled by ZigBee.” *Sensors, 12*(5), 6075-6101.

Welch, B. L. (1947). "The generalization of "Student's" problem when several different population variances are involved". *Biometrika.* 34 (1–2): 28–35.

Ullman, G. L., and Lewis, D. (1998). "Texas DOT vehicle fleet warning light policy research." *The 12th Equipment Management Workshop*, 22.

Zhu, Z., Edara, P., and Sun, C. (2015). "Case study of an alternative merging sign design for temporary traffic control in work zones." *Journal of transportation engineering*, 142(1), 05015005.

APPENDICES

APPENDIX A: SIMULATOR SICKNESS QUESTIONNAIRE (SSQ)

Simulator Sickness Questionnaire

Instructions: Circle how much each symptom below is affecting you right now.

1. General discomfort	None	Slight	Moderate	Severe
2. Fatigue	None	Slight	Moderate	Severe
3. Headache	None	Slight	Moderate	Severe
4. Eye strain	None	Slight	Moderate	Severe
5. Difficult focusing	None	Slight	Moderate	Severe
6. Salivation increasing	None	Slight	Moderate	Severe
7. Sweating	None	Slight	Moderate	Severe
8. Nausea	None	Slight	Moderate	Severe
9. Difficulty concentrating	None	Slight	Moderate	Severe
10. Fullness of the Head	None	Slight	Moderate	Severe
11. Blurred vision	None	Slight	Moderate	Severe
12. Dizziness with eyes open	None	Slight	Moderate	Severe
13. Dizziness with eyes closed	None	Slight	Moderate	Severe
14. *Vertigo	None	Slight	Moderate	Severe
15. **Stomach awareness	None	Slight	Moderate	Severe
16. Burping	None	Slight	Moderate	Severe

* Vertigo is experienced as loss of orientation with respect to vertical upright.
** Stomach awareness is usually used to indicate a feeling of discomfort which is just short of nausea.

APPENDIX B: DATA COLLECTION RESULTS FOR ALL TYPES OF VEHICLES (AFAD FIELD STUDIES)

B.1 Total Data Statistics (Field Study)

Table B.1.1 Total Data Statistics, MOE 1 – MOE 4 (Field Study)

MOE 1 – MOE 4		Total							
		AFAD				Flagger			
		Sedan	Pickup	CV	Total	Sedan	Pickup	CV	Total
Speed at 250 ft. (mi/hr)	Sample Size	57	123	13	193	55	91	9	155
	Mean	23.947	23.211	20.231	23.228	28.273	27.198	23.556	27.368
	SD	6.323	5.553	6.260	5.871	6.404	6.684	4.275	6.527
	Max	39	38	32	39	43	58	30	58
	Min	10	10	10	10	11	16	18	11
	T-test					0.000	0.000	0.183	0.000
Full Stop Location (ft.)	Sample Size	52	102	9	163	49	85	7	141
	Mean	58.942	62.059	62.222	61.074	50.735	49.212	47.143	49.638
	SD	35.207	25.817	31.236	29.259	31.944	15.784	20.178	22.752
	Max	170	170	100	170	220	100	65	220
	Min	10	25	0	0	10	25	10	10
	T-test					0.224	0.000	0.287	0.000
Waiting Time (s)	Sample Size	44	83	7	134	38	71	7	116
	Mean	84.584	62.364	111.991	72.253	98.911	111.602	79.629	105.515
	SD	71.607	74.660	184.048	82.565	79.927	96.231	34.484	88.500
	Max	290.791	548.765	518.518	548.765	367.474	464.508	124.515	464.508
	Min	1.418	2.002	2.976	1.418	2.555	4.721	24.950	2.555
	T-test					0.395	0.000	0.656	0.002
Reaction Time (s) (Based on CMS)	Sample Size	43	83	7	133	38	69	6	113
	Mean	4.500	4.398	4.024	4.412	1.492	1.758	2.171	1.690
	SD	3.179	3.477	1.376	3.290	0.955	0.860	1.008	0.908
	Max	17.491	27.161	6.072	27.161	4.332	4.725	3.433	4.725
	Min	1.193	0.804	2.457	0.804	0.204	0.177	0.612	0.177
	T-test					0.000	0.000	0.020	0.000

Table B.1.2 Total Data Statistics, MOE 5 – MOE 7 (Field Study)

MOE 5 – MOE 7		Total							
		AFAD				Flagger			
		Sedan	Pickup	CV	Total	Sedan	Pickup	CV	Total
Noncompliance	Sample Size	57	123	13	193	55	91	9	155
	Abs. Number	1	1	1	3	2	1	0	3
	Mean	0.018	0.008	0.077	0.016	0.036	0.011	0.000	0.019
	T-test					0.542	0.831	0.419	0.787
1st Following Vehicle Speed at 250 ft. (mph)	Sample Size	25	41	5	71	24	52	6	82
	Mean	22.160	20.195	16.600	20.634	22.667	23.308	22.833	23.085
	SD	5.498	5.105	3.130	5.284	5.639	5.319	5.565	5.371
	Max	31	31	20	31	37	33	30	37
	Min	10	10	14	10	13	13	17	13
	T-test					0.752	0.005	0.054	0.005
Queue Length (veh)	Sample Size	57	123	13	193	55	90	9	154
	Mean	1.825	1.610	2.000	1.699	1.927	2.178	2.111	2.084
	SD	1.167	1.185	2.236	1.272	1.245	1.346	0.928	1.288
	Max	6	8	9	9	6	6	3	6
	Min	1	1	1	1	1	1	1	1
	T-test					0.653	0.001	0.890	0.006

B.2 South Bound Data Statistics (Field Study)

Table B.2.1 South Bound Data Statistics, MOE 1 – MOE 4 (Field Study)

MOE 1 – MOE 4		South							
		AFAD				Flagger			
		Sedan	Pickup	CV	Total	Sedan	Pickup	CV	Total
Speed at 250 ft. (mph)	Sample Size	32	57	8	97	30	47	4	81
	Mean	23.094	22.053	21.875	22.381	26.800	25.149	20.500	25.531
	SD	6.130	4.673	6.978	5.355	5.804	7.587	1.732	6.883
	Max	34	35	32	35	38	58	22	58
	Min	10	13	10	10	17	16	18	16
	T-test					0.018	0.012	0.712	0.001
Full Stop Location (ft.)	Sample Size	32	57	7	96	29	46	3	78
	Mean	58.906	64.912	63.571	62.813	49.862	50.609	31.667	49.603
	SD	41.673	30.713	35.674	34.799	23.760	19.797	22.546	21.464
	Max	170	170	100	170	120	100	55	120
	Min	10	25	0	0	10	25	10	10
	T-test					0.309	0.007	0.197	0.004
Waiting Time (s)	Sample Size	25	41	5	71	21	39	3	63
	Mean	68.055	59.420	30.060	60.393	97.956	129.182	93.266	117.063
	SD	50.508	57.821	33.187	54.178	83.763	117.410	31.777	104.825
	Max	184.962	237.783	85.632	237.783	367.474	464.508	124.515	464.508
	Min	4.010	2.879	2.976	2.879	2.555	4.721	60.987	2.555
	T-test					0.143	0.001	0.038	0.000
Reaction Time (s) (Based on CMS)	Sample Size	25	42	5	72	21	39	2	61
	Mean	4.565	3.821	3.902	4.085	1.491	1.566	1.850	1.551
	SD	4.078	2.865	1.596	3.260	0.765	0.582	0.078	0.636
	Max	17.491	14.932	6.072	17.491	4.194	2.987	1.906	4.194
	Min	1.193	0.804	2.457	0.804	0.365	0.177	1.795	0.177
	T-test					0.001	0.000	0.146	0.000

Table B.2.2 South Bound Data Statistics, MOE 5 – MOE 7 (Field Study)

MOE 5 – MOE 7		South							
		AFAD				Flagger			
		Sedan	Pickup	CV	Total	Sedan	Pickup	CV	Total
Noncompliance	Sample Size	32	57	8	97	30	47	4	81
	Abs Number	0	0	1	0	0	0	0	0
	Mean	0.000	0.000	0.125	0.010	0.000	0.000	0.000	0.000
	T-test					\	\	0.506	0.356
1st Following Vehicle Speed at 250 ft. (mph)	Sample Size	12	17	2	31	14	31	3	48
	Mean	20.833	17.412	20.000	18.903	20.714	22.290	22.333	21.833
	SD	3.538	5.432	0.000	4.812	3.451	5.503	6.807	5.012
	Max	27	25	20	27	28	33	30	33
	Min	15	10	20	10	16	13	17	13
	T-test					0.932	0.005	0.677	0.012
Queue Length (veh)	Sample Size	32	57	8	97	30	47	4	81
	Mean	1.719	1.509	1.250	1.557	1.967	2.319	2.250	2.185
	SD	1.170	0.966	0.707	1.020	1.299	1.337	0.957	1.305
	Max	6	5	3	6	6	6	3	6
	Min	1	1	1	1	1	1	1	1
	T-test					0.432	0.001	0.066	0.000

B.3: North Bound Data Statistics (Field Study)

Table B.3.1 North Bound Data Statistics, MOE 1 – MOE 4

MOE 1 – MOE 4		North							
		AFAD				Flagger			
		Sedan	Pickup	CV	Total	Sedan	Pickup	CV	Total
Speed at 250 ft. (mile/hr)	Sample Size	25	66	5	96	25	44	5	74
	Mean	25.040	24.212	17.600	24.083	30.040	29.386	26.000	29.378
	SD	6.522	6.073	4.278	6.262	6.755	4.736	4.183	5.486
	Max	39	38	23	39	43	38	30	43
	Min	15	10	12	10	11	18	20	11
	T-test					0.011	0.000	0.014	0.000
Full Stop Location (ft.)	Sample Size	20	45	2	67	20	39	4	63
	Mean	59.000	58.444	57.500	58.582	52.000	47.564	58.750	49.683
	SD	22.219	17.478	10.607	18.644	41.751	9.023	7.500	24.427
	Max	120	120	65	120	220	70	65	220
	Min	35	40	50	35	30	25	50	25
	T-test					0.512	0.001	0.872	0.021
Waiting Time (s)	Sample Size	19	42	2	63	17	32	4	53
	Mean	106.334	65.238	316.817	85.619	100.090	90.177	69.401	91.789
	SD	89.269	88.710	285.249	104.724	77.460	56.122	37.144	62.179
	Max	290.791	548.765	518.518	548.765	297.378	229.726	110.055	297.378
	Min	1.418	2.002	115.115	1.418	5.024	5.393	24.950	5.024
	T-test					0.825	0.168	0.122	0.707
Reaction Time (s) (Based on CMS)	Sample Size	18	41	2	61	18	30	4	52
	Mean	4.410	4.990	4.330	4.797	1.493	2.007	2.332	1.854
	SD	1.214	3.958	0.955	3.310	1.154	1.085	1.260	1.133
	Max	7.691	27.161	5.005	27.161	4.332	4.725	3.433	4.725
	Min	2.520	1.718	3.654	1.718	0.204	0.734	0.612	0.204
	T-test					0.000	0.000	0.125	0.000

Table B.3.2 North Bound Data Statistics, MOE 5 – MOE 7 (Field Study)

MOE 5 – MOE 7		North							
		AFAD				Flagger			
		Sedan	Pickup	CV	Total	Sedan	Pickup	CV	Total
Noncompliance	Sample Size	25	66	5	96	25	44	5	74
	Abs Number	1	1	0	2	2	1	0	3
	Mean	0.040	0.015	0.000	0.021	0.080	0.023	0.000	0.041
	T-test					0.561	0.773	\	0.454
1st Following Vehicle Speed at 250 ft. (mph)	Sample Size	13	24	3	40	10	21	3	34
	Mean	23.385	22.167	14.333	21.975	25.400	24.810	23.333	24.853
	SD	6.752	3.875	0.577	5.299	7.058	4.771	5.508	5.434
	Max	31	31	15	31	37	33	27	37
	Min	10	14	14	10	13	18	17	13
	T-test					0.494	0.047	0.048	0.024
Queue Length (veh)	Sample Size	25	66	5	96	25	43	5	73
	Mean	1.960	1.697	3.200	1.844	1.880	2.023	2.000	1.973
	SD	1.172	1.347	3.347	1.475	1.201	1.354	1.000	1.269
	Max	5	8	9	9	4	5	3	5
	Min	1	1	1	1	1	1	1	1
	T-test					0.813	0.220	0.464	0.551

APPENDIX C: AFAD FIELD SURVEY QUESTIONS

Date ______________________

Work Zone Signage Survey

Note: To complete this survey online using a computer or mobile device, please visit https://goo.gl/BM40Ju or scan the QR code below.

Proper communication of work zone information is critical for the safe movement of traffic through work zones. Please provide us with your perspective on the following communication alternatives.

Please refer to the device shown below in Figure 1.

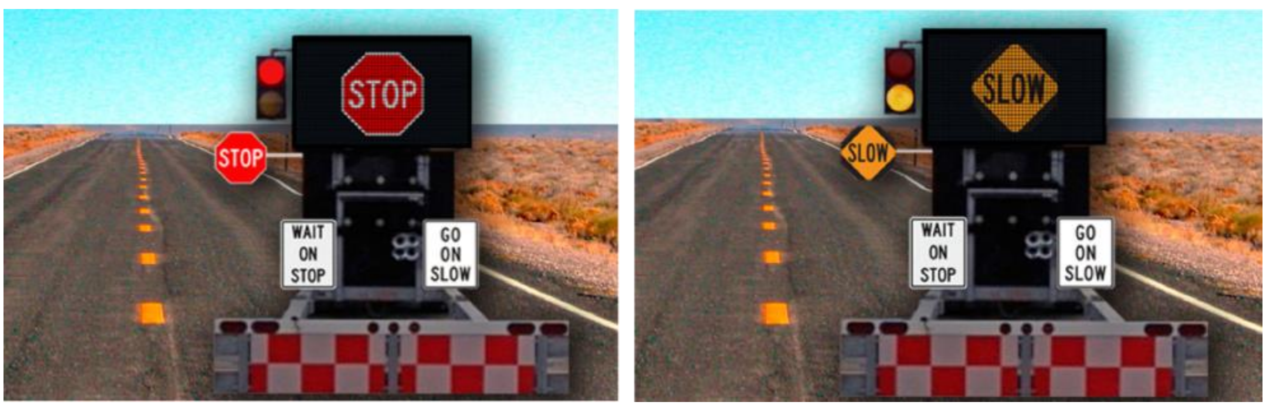


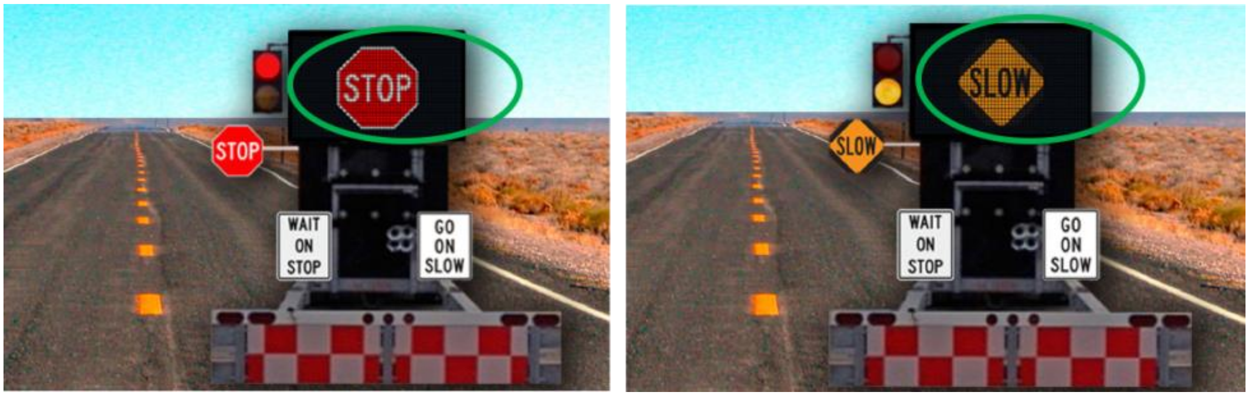
Figure 1

1. **What is the meaning of the device shown in Figure 1?**
   a. Narrow lanes ahead – reduce speed.
   b. Wait if "stop" indicated, proceed if "slow" indicated.
   c. The device makes no sense.

2. **Please rate the effectiveness of the device shown in Figure 1.**
   [ ] Very Effective [ ] Effective [ ] Neutral [ ] Ineffective [ ] Very Ineffective

3. **Please check any reasons for your rating on the device shown in Figure 1.**
   [ ] Clarity [ ] Visibility [ ] Safety [ ] Efficiency
   [ ] Other

____



4. **The message board on the device in Figure 1 (circled in green) was helpful in complementing the instructions provided by the stop/slow paddle.**
   [ ] Strongly agree [ ] Agree [ ] Neutral [ ] Disagree [ ] Strongly disagree

5. **I have encountered the device shown in Figure 1 before.**
   [ ] Yes [ ] No [ ]

6. **Please enter any additional comments you may have regarding the device shown in Figure 1.**

Figure 2

7. **What is the meaning of the signage shown in Figure 2?**
   a. Narrow lanes ahead – reduce speed.
   b. Wait if "stop" indicated, proceed if "slow" indicated.
   c. The signage makes no sense.

8. **Please rate the effectiveness of the signage shown in Figure 2.**
   [ ] Very Effective [ ] Effective [ ] Neutral [ ] Ineffective [ ] Very Ineffective

9. **Please check any reasons for your rating on the signage shown in Figure 2.**
   [ ] Clarity [ ] Visibility [ ] Safety [ ] Efficiency [ ] Other

---

10. I have encountered the signage shown in Figure 2 before.
[ ] Yes [ ] No

11. Please enter any additional comments you may have regarding the signage shown in Figure 2.

____

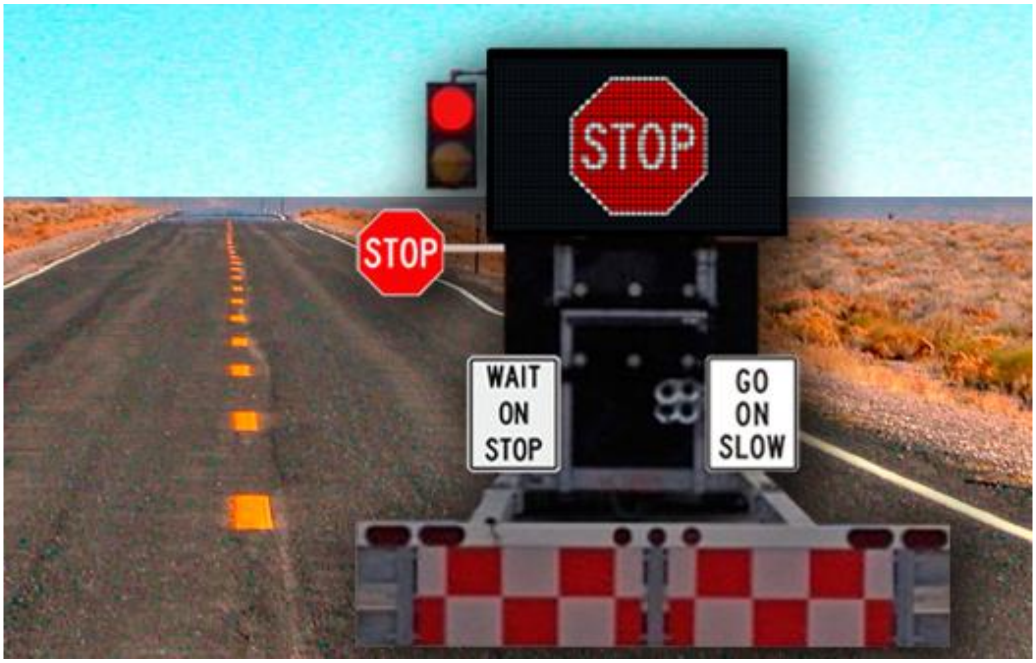


Figure 1

Figure 2

12. Please indicate your preference.
[ ] Figure 1 much more [ ] Figure 1 more [ ] Neutral [ ] Figure 2 more [ ] Figure 2 much more

Please answer the demographic questions below.
13. Age range
[ ] 16-25 [ ] 26-40 [ ] 41-55 [ ] 56-70 [ ] 71-95

14. Gender
[ ] Male [ ] Female

15. My Residency
[ ] Urban [ ] Rural

16. My Regular Vehicle Type
[ ] Passenger Car [ ] Vehicle towing trailer [ ] Delivery/Moving Truck
[ ] Tractor trailer truck [ ] Bus

Please contact Mr. Henry Brown (**brownhen@missouri.edu**) for additional comments, concerns or information on this survey. Thank you for completing this survey! We greatly appreciate your time!

APPENDIX D: AFAD FIELD SURVEY RESULTS BY DIFFERENT GROUPS

D.1 Results by Age (Field Survey)

Table D.1.1 Age Range vs. Effectiveness of AFAD (Figure 1 in Field Survey)

Age Ranges	Effectiveness	Count
16-25	Very Effective	1
26-40	Very Effective	3
	Effective	1
41-55	Very Effective	4
	Effective	3
	Neutral	1
56-70	Very Effective	9
	Effective	2
	Ineffective	1
	Very Ineffective	2
71-95	Very Effective	9
	Effective	3
	Very Ineffective	1

Table D.1.2 Age Range vs. Reasons of Rating (Field Survey)

Age Ranges	Factor	Count	Age Ranges	Factor	Count
16-25	Clarity	1	56-70	Clarity	12
	Visibility	1		Visibility	11
	Safety	1		Safety	10
	Efficiency	1		Efficiency	7
	Other	0		Other	0
26-40	Clarity	4	71-95	Clarity	9
	Visibility	5		Visibility	12
	Safety	5		Safety	10
	Efficiency	3		Efficiency	6
	Other	2		Other	0
41-55	Clarity	5	Total sample size = 41		
	Visibility	7			
	Safety	6			
	Efficiency	3			
	Other	3			

Table D.1.3 Age Range vs. Helpfulness of CMS (Figure 1 in Field Survey)

Age Ranges	Categories	Count
16-25	Strongly Agree	1
26-40	Strongly Agree	3
	Agree	2
41-55	Strongly Agree	5
	Agree	1
	Neutral	1
	Strongly Disagree	1
56-70	Strongly Agree	9
	Agree	3
	Neutral	2
71-95	Strongly Agree	5
	Agree	8

Table D.1.4 Age Range vs. Encountered Signage Before (Field Survey)

Age Ranges	Yes or No	Count
16-25	Yes	1
	No	0
26-40	Yes	4
	No	1
41-55	Yes	4
	No	4
56-70	Yes	4
	No	10
71-95	Yes	6
	No	7

Table D.1.5 Age Range vs. Effectiveness of Flagger (Figure 2 in Field Survey)

Age Ranges	Effectiveness	Count
16-25	Very Effective	1
26-40	Effective	5
41-55	Effective	8
56-70	Very Effective	3
	Effective	9
	Ineffective	1
	Very Ineffective	1
71-95	Very Effective	3
	Effective	9
	Neutral	1

Table D.1.6 Age Range vs. Reasons of Rating (Field Survey)

Age Ranges	Factor	Count
16-25	Clarity	1
	Visibility	1
	Safety	1
	Efficiency	1
	Other	0
26-40	Clarity	2
	Visibility	3
	Safety	2
	Efficiency	2
	Other	0
41-55	Clarity	4
	Visibility	3
	Safety	1
	Efficiency	3
	Other	2
56-70	Clarity	7
	Visibility	6
	Safety	8
	Efficiency	2
	Other	1
71-95	Clarity	9
	Visibility	8
	Safety	10
	Efficiency	6
	Other	1

Table D.1.7 Age Range vs. Preference (Field Survey)

Age Ranges	Preference	Count
16-25	Neutral	1
26-40	AFAD Much More	3
	AFAD More	2
41-55	AFAD Much More	6
	AFAD More	1
	Flagger More	1
56-70	AFAD Much More	8
	AFAD More	2
	Neutral	1
	Flagger More	2
71-95	AFAD Much More	4
	AFAD More	5
	Neutral	2
	Flagger More	2

D.2 Results by Gender (Field Survey)

Table D.2.1 Gender vs. Effectiveness of AFAD

Gender	Effectiveness	Count
Male	Very Effective	16
	Effective	3
	Neutral	1
	Ineffective	1
	Very Ineffective	1
Female	Very Effective	11
	Effective	6
	Ineffective	0
	Very Ineffective	2

Table D.2.2 Gender vs. Reasons of Rating (Field Survey)

Gender	Factor	Count
Male	Clarity	14
	Visibility	19
	Safety	15
	Efficiency	10
	Other	2
Female	Clarity	17
	Visibility	17
	Safety	15
	Efficiency	10
	Other	3

Table D.2.3 Gender vs. Helpfulness of CMS (Field Survey)

Gender	Category	Count
Male	Strongly Agree	12
	Agree	8
	Neutral	1
	Strongly Disagree	1
Female	Strongly Agree	11
	Agree	6
	Neutral	2

Table D.2.4 Gender vs. Encountered Stop Control Before (Field Survey)

Gender	Yes or No	Count
Male	Yes	10
	No	12
Female	Yes	9
	No	10

Table D.2.5 Gender vs. Effectiveness of Flagger (Field Survey)

Gender	Effectiveness	Count
Male	Very effective	4
	Effective	17
	Ineffective	1
Female	Very effective	3
	Effective	14
	Neutral	1
	Very Ineffective	1

Table D.2.6 Gender vs. Reasons of Rating (Field Survey)

Gender	Factor	Count
Male	Clarity	10
	Visibility	12
	Safety	11
	Efficiency	9
	Other	1
Female	Clarity	11
	Visibility	11
	Safety	9
	Efficiency	4
	Other	4

Table D.2.7 Gender vs. Preference (Field Survey)

Gender	Preference	Count
Male	AFAD Much More	12
	AFAD More	4
	Neutral	2
	Flagger More	4
Female	AFAD Much More	9
	AFAD More	6
	Neutral	2
	Flagger More	1

D.3 Survey Results by Residency (Field Survey)

Table D.3.1 Residency vs. Effectiveness of AFAD (Field Survey)

Residency	Effectiveness	Count
Urban	Very Effective	4
	Effective	1
	Very Ineffective	1
Rural	Very Effective	23
	Effective	8
	Neutral	1
	Ineffective	1
	Very Ineffective	2

Table D.3.2 Residency vs. Reasons of Rating (Field Survey)

Residency	Factor	Count
Urban	Clarity	6
	Visibility	4
	Safety	4
	Efficiency	3
	Other	0
Rural	Clarity	25
	Visibility	32
	Safety	26
	Efficiency	17
	Other	5

Table D.3.3 Residency vs. Helpfulness of CMS (Field Survey)

Residency	Categories	Count
Urban	Strongly Agree	3
	Agree	3
Rural	Strongly Agree	20
	Agree	11
	Neutral	3
	Strongly Disagree	1

Table D.3.4 Residency vs. Encountered Stop Control Before (Field Survey)

Residency	Yes or No	Count
Urban	Yes	3
	No	3
Rural	Yes	16
	No	19

Table D.3.5 Residency vs. Effectiveness of Flagger (Field Survey)

Residency	Effectiveness	Count
Urban	Very Effective	2
	Effective	3
	Ineffective	1
Rural	Very Effective	5
	Effective	28
	Neutral	1
	Very Ineffective	1

Table D.3.6 Residency vs. Reasons of Rating (Field Survey)

Age Ranges	Factor	Count
Urban	Clarity	3
	Visibility	6
	Safety	3
	Efficiency	2
	Other	0
Rural	Clarity	18
	Visibility	17
	Safety	17
	Efficiency	11
	Other	5

Table D.3.7 Residency vs. Preference (Field Survey)

Residency	Preference	Count
Urban	AFAD Much More	2
	AFAD More	2
	Neutral	2
Rural	AFAD Much More	19
	AFAD More	8
	Neutral	2
	Flagger More	5

APPENDIX E: AFAD SIMULATOR RESULTS

Table E.0.1 Simulator Results

	MOE 1 Approach Speed (mph)							
	Mean	SD	Diff	Cohen's d	Confidence Level	Diff	Cohen's d	Confidence Level
Flagger	34.79	13.83	Baseline			n/a		
MoDOT AFAD	26.34	11.63	-8.44	-0.66	> 99.9%*	Baseline		
AFAD with alternative sign	25.98	10.30	-8.80	-0.72	> 99.9%*	-0.36	0.03	28.2%
AFAD without CMS	26.87	11.07	-7.91	-0.63	> 99.9%*	0.53	0.05	33.7%
	MOE 2 Full Stop Distance (feet)							
	Mean	SD	Diff	Cohen's d	Confidence Level	Diff	Cohen's d	Confidence Level
Flagger	53.09	36.03	Baseline			n/a		
MoDOT AFAD	97.55	49.93	44.46	1.02	> 99.9%*	Baseline		
AFAD with alternative sign	90.67	48.69	37.58	0.88	> 99.9%*	-6.88	-0.14	89.6%
AFAD without CMS	74.20	28.20	21.11	0.65	> 99.9%*	-23.35	-0.58	> 99.9%*
	MOE 3 Reaction Time (seconds)							
	Mean	SD	Diff	Cohen's d	Confidence Level	Diff	Cohen's d	Confidence Level
Flagger	2.05	1.14	Baseline			n/a		
MoDOT AFAD	1.93	1.99	-0.12	-0.03	70.3%	Baseline		
AFAD with alternative sign	1.60	1.86	-0.45	-0.20	94.5%	-0.32	-0.13	82.9%
AFAD without CMS	1.23	1.84	-0.82	-0.56	99.8%*	-0.70	-0.42	98.8%*
	MOE 4 Noncompliance Rate							
	Mean	SD	Diff	Cohen's d	Confidence Level	Diff	Cohen's d	Confidence Level
Flagger	0.14	0.35	Baseline			n/a		
MoDOT AFAD	0.00	0.00	-0.14	N. A	99.8%*	Baseline		
AFAD with alternative sign	0.00	0.00	-0.14	N. A	99.8%*	0.00	n/a	n/a
AFAD without CMS	0.05	0.21	-0.09	-0.32	94.3%	0.05	n/a	91.7%

* indicates significance at 99% confidence level

	MOE 5 First Brake Location (feet)							
	Mean	SD	Diff	Cohen’s d	Confidence Level	Diff	Cohen’s d	Confidence Level
Flagger	274.02	120.51	Baseline			n/a		
MoDOT AFAD	332.19	108.55	58.17	0.51	99.5%*	Baseline		
AFAD with alternative sign	334.95	112.08	60.94	0.52	99.5%*	2.77	-0.03	13.9%
AFAD without CMS	320.30	106.09	46.29	0.41	99.5%*	-11.89	-0.11	46.6%

* indicates significance at 99% confidence level

APPENDIX F: AFAD SIMULATOR SURVEY QUESTIONS

Date ________________________

AFAD Simulator Survey

Proper communication of work zone information is critical for the safe movement of traffic through work zones.

Please answer the following questions.

Figure 1a

1. **What is the meaning of Figure 1a?**
   a. Wait if STOP sign indicated.
   b. A regular STOP sign, make a full stop and go.
   c. The device makes no sense.

Figure 1b

2. **What is the meaning of Figure 1b?**
   a. Wait if STOP sign indicated.
   b. A regular STOP sign, make a full stop and go.
   c. This is a traffic signal, stop on the red light.
   d. The device makes no sense.

Figure 1c

3. **What is the meaning of Figure 1c?**
    a. Wait if STOP sign indicated.
    b. A regular STOP sign, make a full stop and go.
    c. This is a traffic signal, stop on the red light.
    d. The device makes no sense.

Figure 1d

4. **What is the meaning of Figure 1d?**
    a. Wait if STOP sign indicated.
    b. A regular STOP sign, make a full stop and go.
    c. This is a traffic signal, stop on the red light.
    d. The device makes no sense.


Figure 1a

Figure 1b


Figure 1c


Figure 1d

5. **Please rank your preference from [1] being most preferred to [4] being least preferred.**

[ ] Figure 1a [ ] Figure 1b [ ] Figure 1c [ ] Figure 1d

6. **Please rate all designs from a scale of 1 (lowest) to 10 (highest) with respect to the following attributes:**

	Figure 1a 	**Figure 1b** 	**Figure 1c** 	**Figure 1d** 
Clarity				
Visibility				
Safety				
Efficiency				
Comments				

Figure 2

7. **To what extent do you believe the message board in Figure 2, as circled in green, is necessary?**

[ ] Strongly agree [ ] Agree [ ] Neutral [ ] Disagree [ ] Strongly disagree

8. **I felt like I was actually there on the highway.**

[ ] Strongly agree [ ] Agree [ ] Neutral [ ] Disagree [ ] Strongly disagree

9. **I felt like I could drive around freely.**

[ ] Strongly agree [ ] Agree [ ] Neutral [ ] Disagree [ ] Strongly disagree

10. **Did any issues arise during the use of the Automated Flagger Assistance Devices (AFAD) simulator?**

[ ] Yes [ ] No
If yes, please explain the issue(s) that you experienced:
____
____
____

Please answer the demographic questions below.

11. **Age range**

[ ] 16-25 [ ] 26-40 [ ] 41-55 [ ] 56-70 [ ] 71-95

12. **Gender**

[ ] Male [ ] Female

13. **My Residency**

[ ] Urban [ ] Rural

14. **My Regular Vehicle Type**

[ ] Passenger Car [ ] Vehicle towing trailer [ ] Delivery/Moving Truck
[ ] Tractor trailer truck [ ] Bus

15. **Please enter any additional comments you may have regarding this study.**

____
____
____

Please contact Mr. Henry Brown (brownhen@missouri.edu) for additional comments, concerns or information on this survey. Thank you for completing this survey! We greatly appreciate your time!

APPENDIX G: ONE WAY ANOVA RESULTS FOR CLARITY, VISIBILITY, SAFETY, AND EFFICIENCY (AFAD SIMULATOR SURVEY)

This appendix shows the one-way ANOVA test results of clarity, visibility, safety and efficiency rated by simulator participants. MoDOT AFAD was regarded as the base factor, and human flagger, AFAD with alternative sign, and AFAD without CMS were the alternatives. The grey dots plotted in the figures represent the original individual ratings while the blue dots indicate the mean value of the same corresponding factor. For example, for clarity of flagger vs. MoDOT AFAD (Figure G.1.1), Point (10, 1) means a participant rated "10" for MoDOT AFAD clarity, and rate "1" for human flagger clarity. Each point represented one paired score rated by one participant.

G.1 Clarity

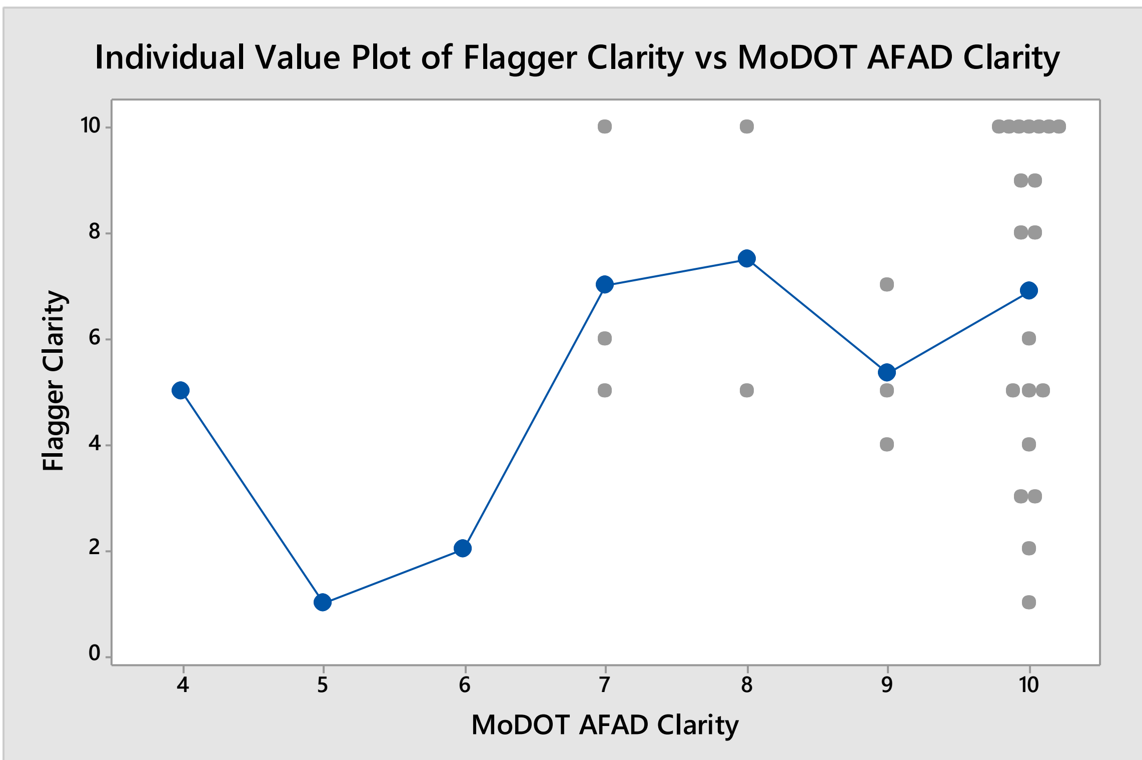


Figure G.1.1 Clarity: Flagger vs. MoDOT AFAD

MoDOT AFAD Clarity	N	Mean	StDev	95% CI
4	1	5.000	*	(-1.095, 11.095)
5	1	1.000	*	(-5.095, 7.095)
6	1	2.000	*	(-4.095, 8.095)
7	4	7.00	2.16	(3.95, 10.05)
8	2	7.50	3.54	(3.19, 11.81)
9	3	5.333	1.528	(1.814, 8.852)
10	20	6.900	3.144	(5.537, 8.263)

Pooled StDev = 2.95950

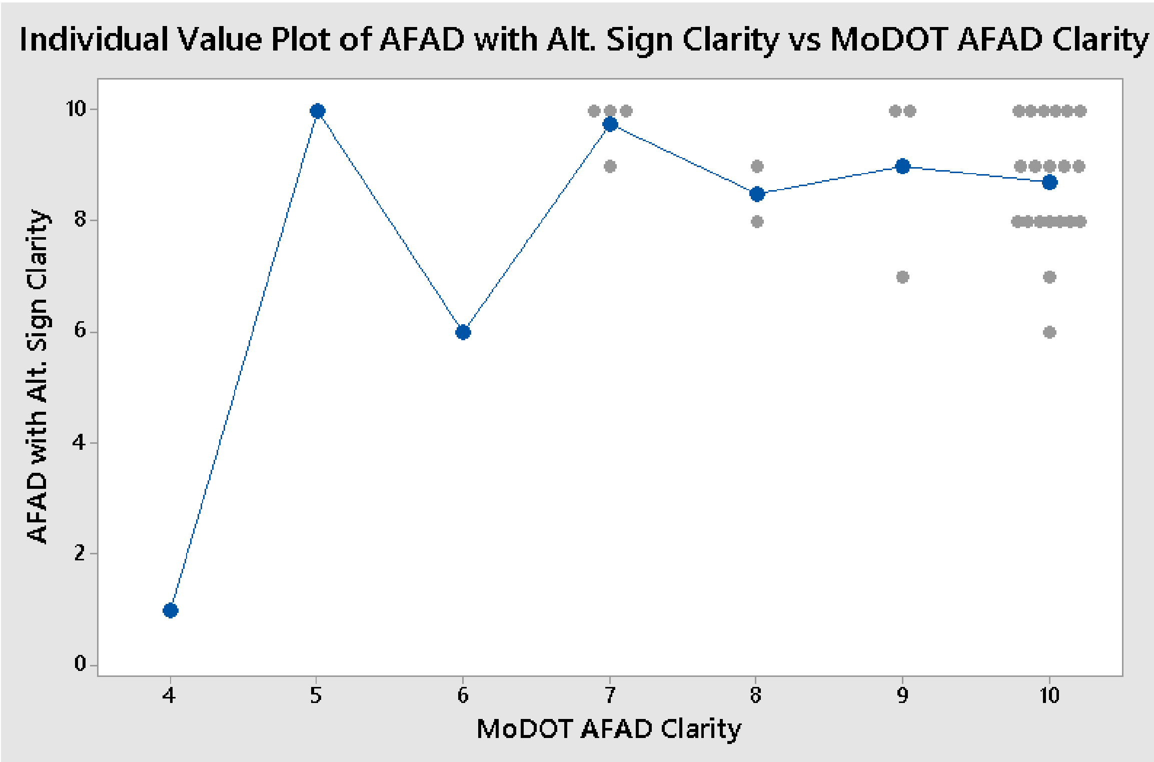


Figure G.1.2 Clarity: AFAD with Alternative Sign vs. MoDOT AFAD

MoDOT AFAD Clarity	N	Mean	StDev	95% CI
4	1	1.000	*	(-1.310, 3.310)
5	1	10.00	*	(7.69, 12.31)
6	1	6.000	*	(3.690, 8.310)
7	4	9.750	0.500	(8.595, 10.905)
8	2	8.500	0.707	(6.867, 10.133)
9	3	9.00	1.73	(7.67, 10.33)
10	20	8.700	1.129	(8.183, 9.217)

Pooled StDev = 1.12161

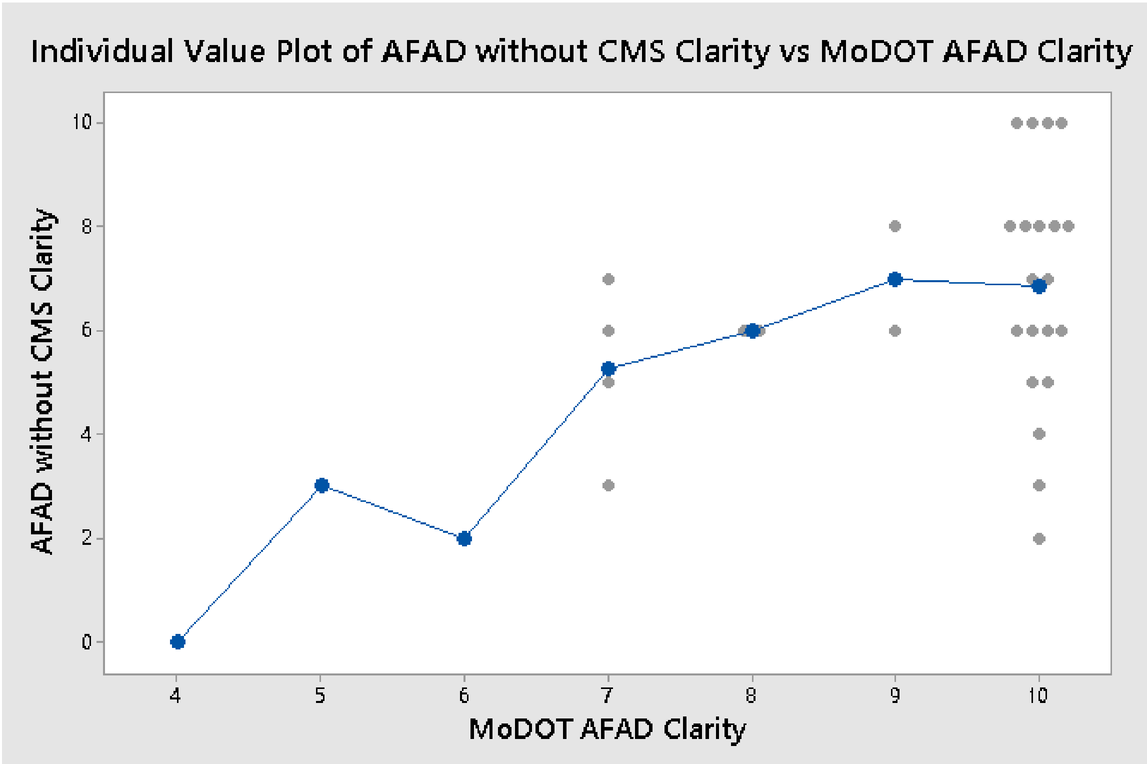


Figure G.1.3 Clarity: AFAD without CMS vs. MoDOT AFAD

MoDOT AFAD Clarity	N	Mean	StDev	95% CI
4	1	0.000000	*	(-4.384448, 4.384448)
5	1	3.000	*	(-1.384, 7.384)
6	1	2.000	*	(-2.384, 6.384)
7	4	5.250	1.708	(3.058, 7.442)
8	2	6.000	0.000	(2.900, 9.100)
9	3	7.000	1.000	(4.469, 9.531)
10	20	6.850	2.323	(5.870, 7.830)

Pooled StDev = 2.12885

G.2 Visibility

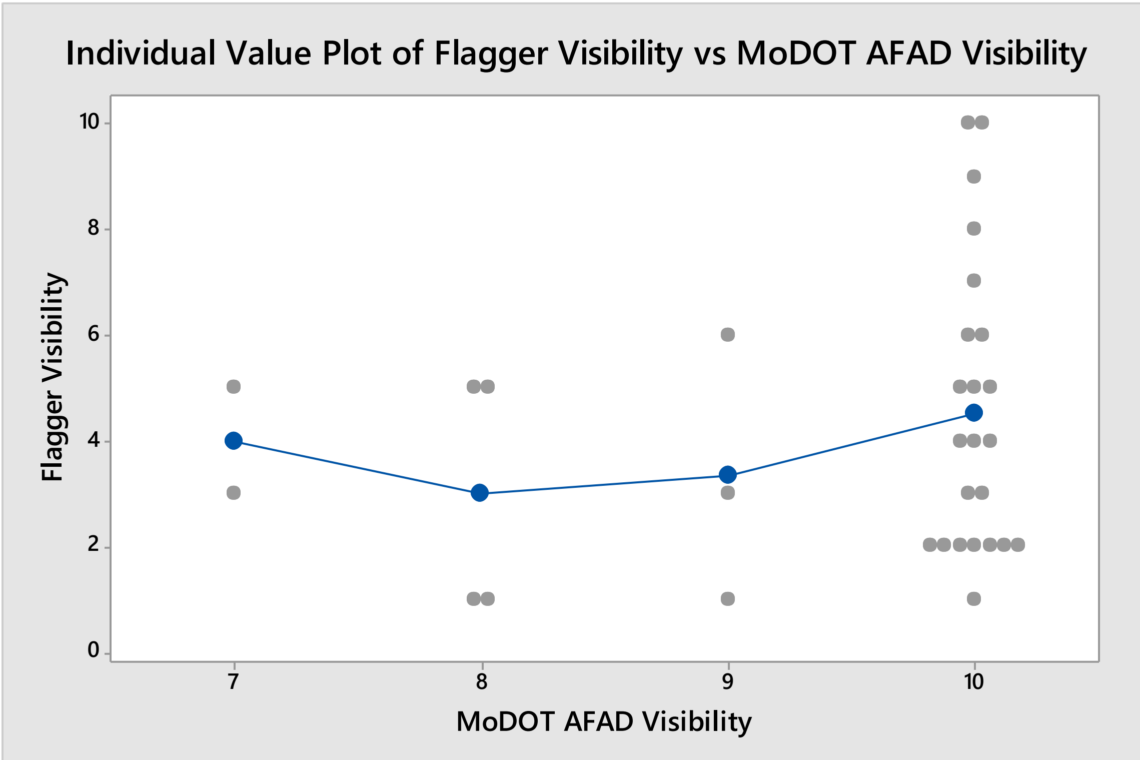


Figure G.2.1 Visibility: Flagger vs. MoDOT AFAD

MoDOT AFAD Visibility	N	Mean	StDev	95% CI
7	2	4.00	1.41	(0.16, 7.84)
8	4	3.00	2.31	(0.29, 5.71)
9	3	3.33	2.52	(0.20, 6.47)
10	23	4.522	2.745	(3.391, 5.653)

Pooled StDev = 2.64849

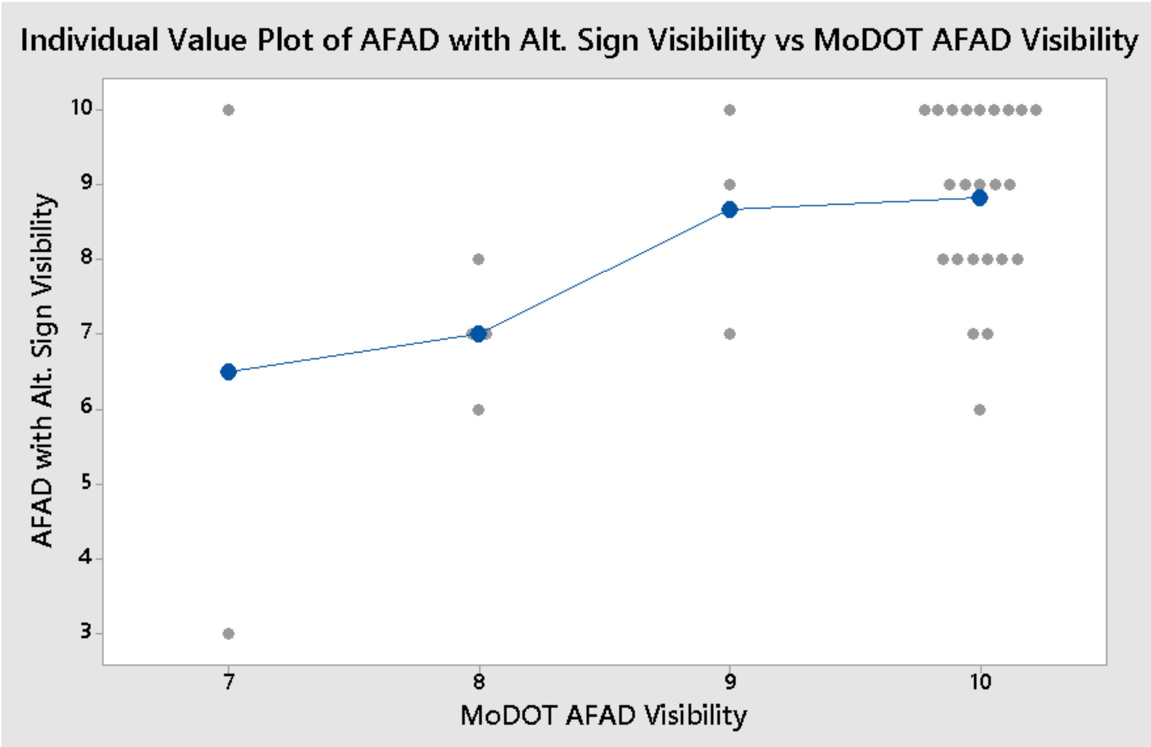


Figure G.2.2 Visibility: AFAD with Alternative Sign vs. MoDOT AFAD

MoDOT AFAD Visibility	N	Mean	StDev	95% CI
7	2	6.50	4.95	(4.34, 8.66)
8	4	7.000	0.816	(5.470, 8.530)
9	3	8.667	1.528	(6.900, 10.433)
10	23	8.826	1.193	(8.188, 9.464)

Pooled StDev = 1.49369

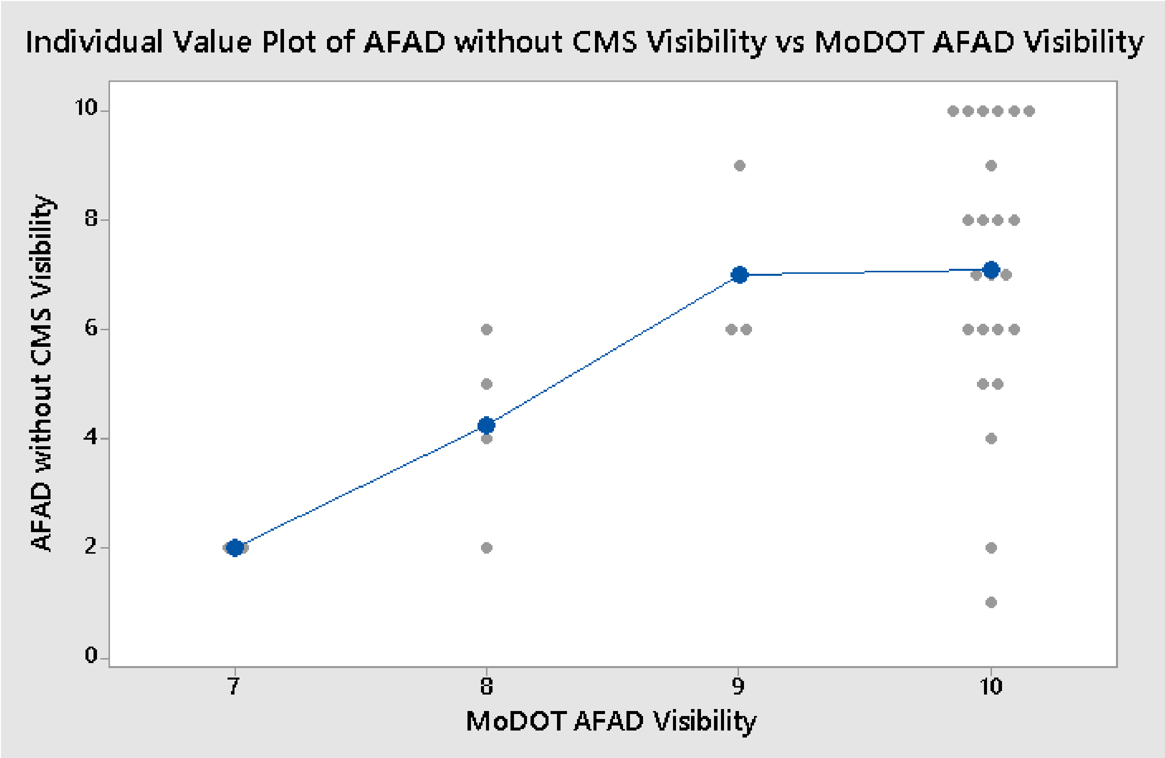


Figure G.2.3 Visibility: AFAD without CMS vs. MoDOT AFAD

MoDOT AFAD Visibility	N	Mean	StDev	95% CI
7	2	2.000	0.000	(-1.447, 5.447)
8	4	4.250	1.708	(1.813, 6.687)
9	3	7.00	1.73	(4.19, 9.81)
10	23	7.087	2.557	(6.070, 8.103)

Pooled StDev = 2.37980

G.3 Safety

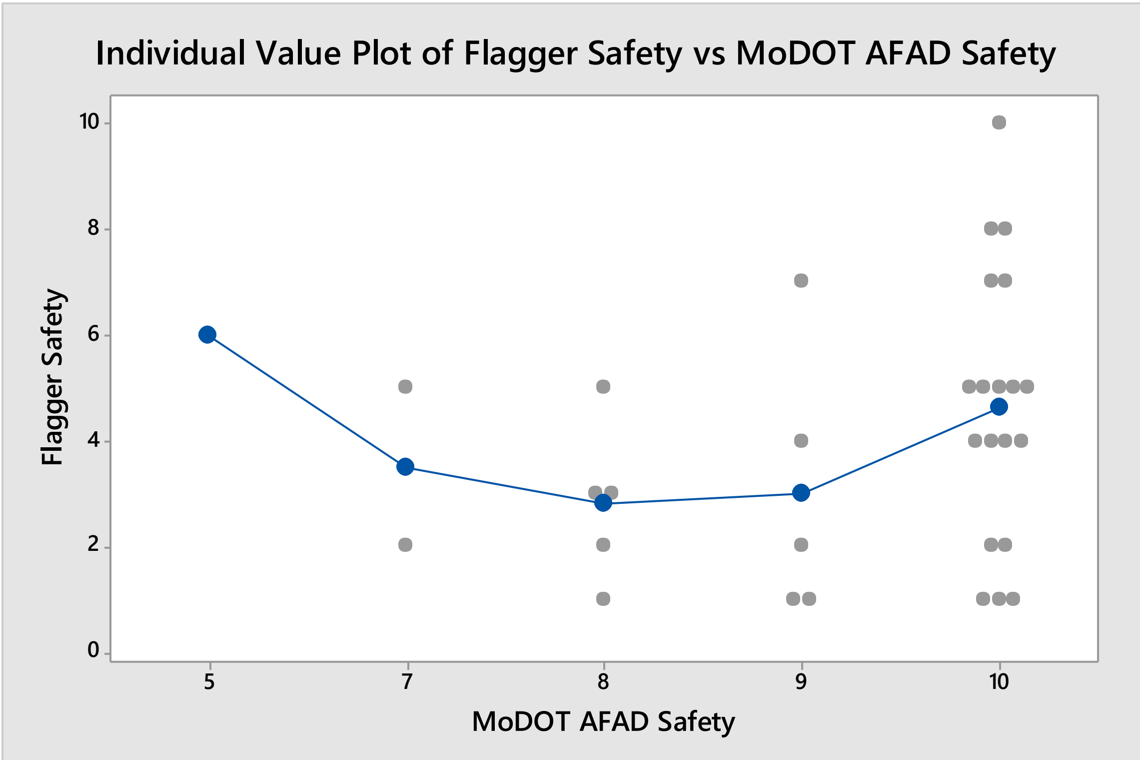


Figure G.3.1 Safety: Flagger vs. MoDOT AFAD

MoDOT AFAD Safety	N	Mean	StDev	95% CI
5	1	6.000	*	(1.041, 10.959)
7	2	3.50	2.12	(-0.01, 7.01)
8	5	2.800	1.483	(0.582, 5.018)
9	5	3.00	2.55	(0.78, 5.22)
10	19	4.632	2.565	(3.494, 5.769)

Pooled StDev = 2.41692

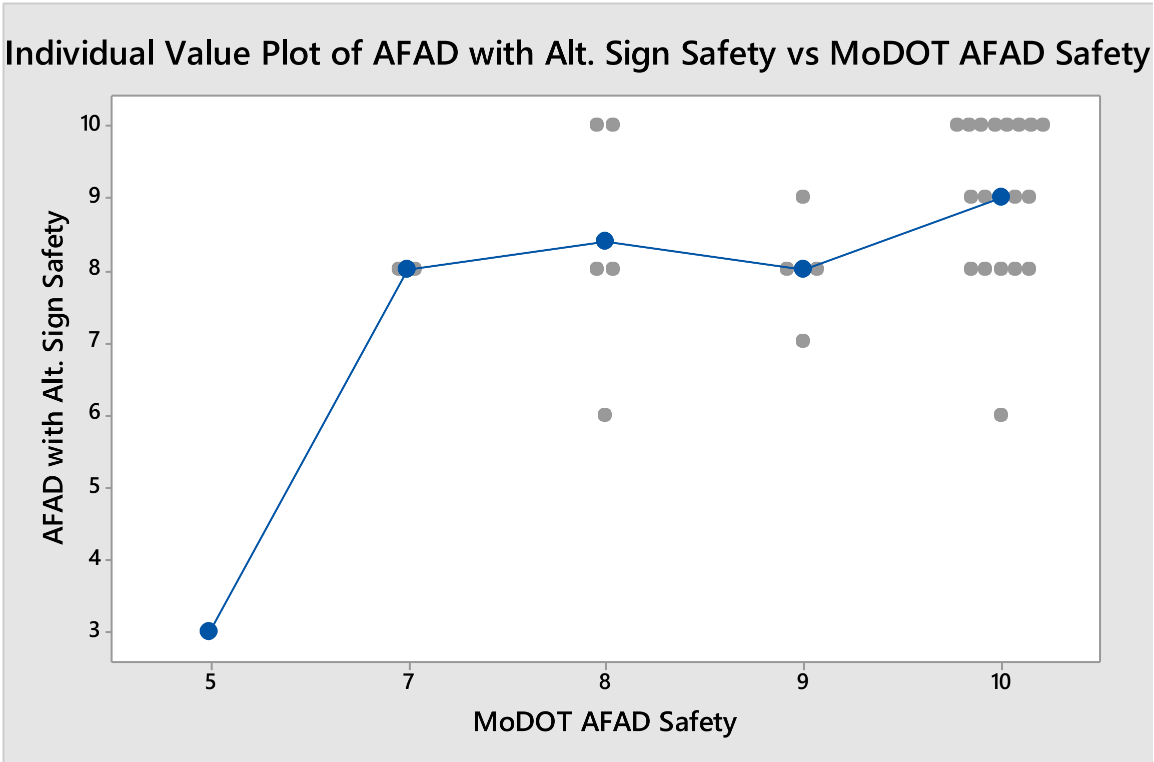


Figure G.3.2 Safety: AFAD with Alternative Sign vs. MoDOT AFAD

```
MoDOT
AFAD
Safety     N   Mean  StDev       95% CI
5          1  3.000      *  (0.657, 5.343)
7          2  8.000  0.000  (6.343, 9.657)
8          5  8.400  1.673  (7.352, 9.448)
9          5  8.000  0.707  (6.952, 9.048)
10        19  9.000  1.106  (8.463, 9.537)

Pooled StDev = 1.14180
```

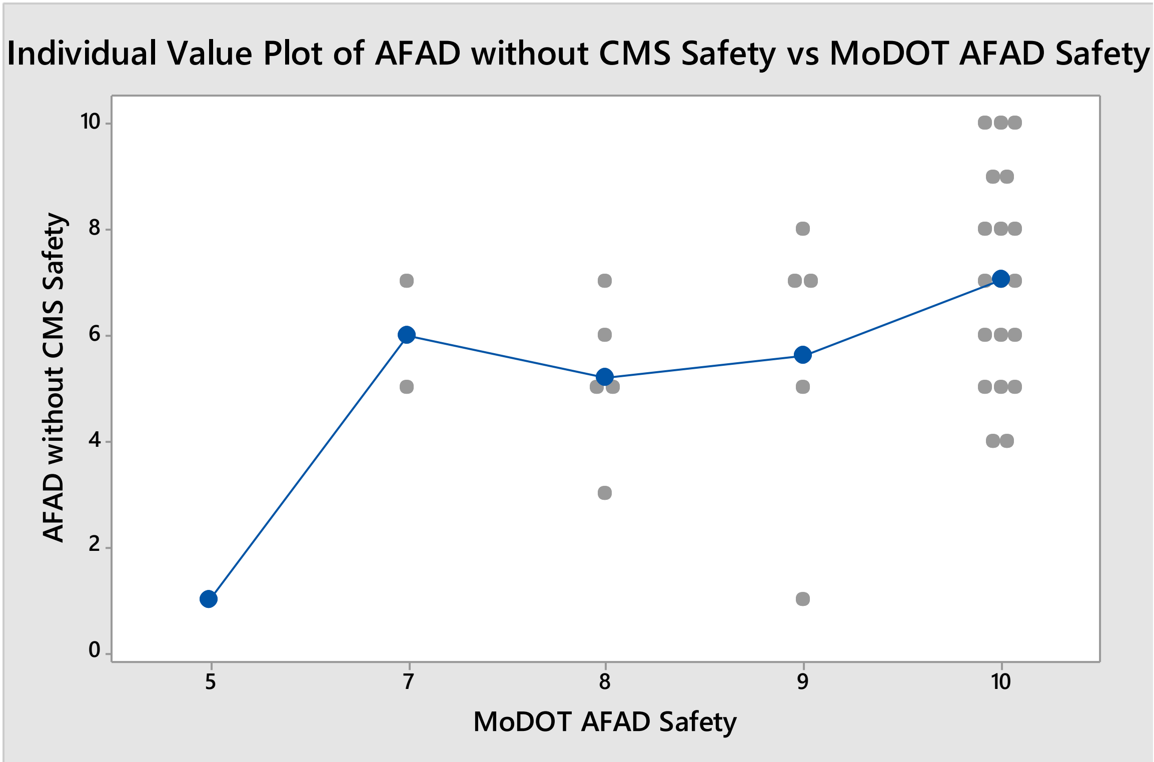


Figure G.3.3 Safety: AFAD without CMS vs. MoDOT AFAD

MoDOT AFAD Safety	N	Mean	StDev	95% CI
5	1	1.000	*	(-3.197, 5.197)
7	2	6.00	1.41	(3.03, 8.97)
8	5	5.200	1.483	(3.323, 7.077)
9	5	5.60	2.79	(3.72, 7.48)
10	19	7.053	1.985	(6.090, 8.015)

Pooled StDev = 2.04530

G.4 Efficiency

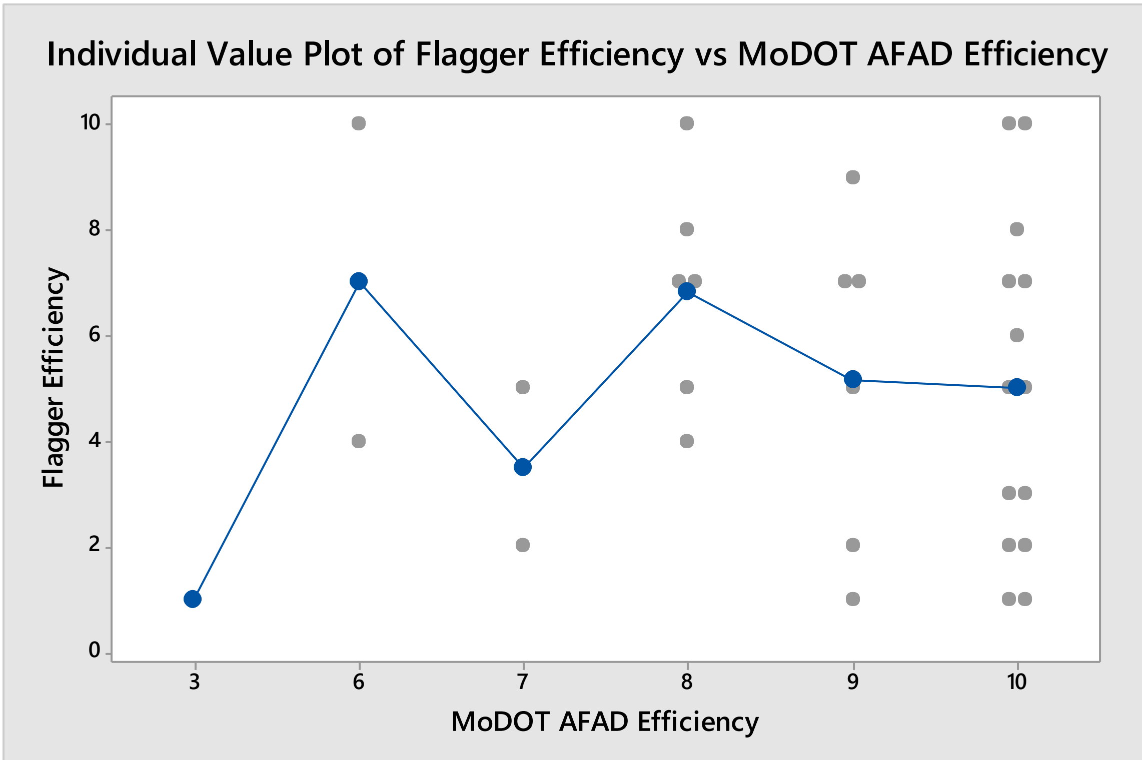


Figure G.4.1 Efficiency: Flagger vs. MoDOT AFAD

MoDOT AFAD Efficiency	N	Mean	StDev	95% CI
3	1	1.000	*	(-5.112, 7.112)
6	2	7.00	4.24	(2.68, 11.32)
7	2	3.50	2.12	(-0.82, 7.82)
8	6	6.833	2.137	(4.338, 9.329)
9	6	5.17	3.13	(2.67, 7.66)
10	14	5.000	3.113	(3.367, 6.633)

Pooled StDev = 2.96760

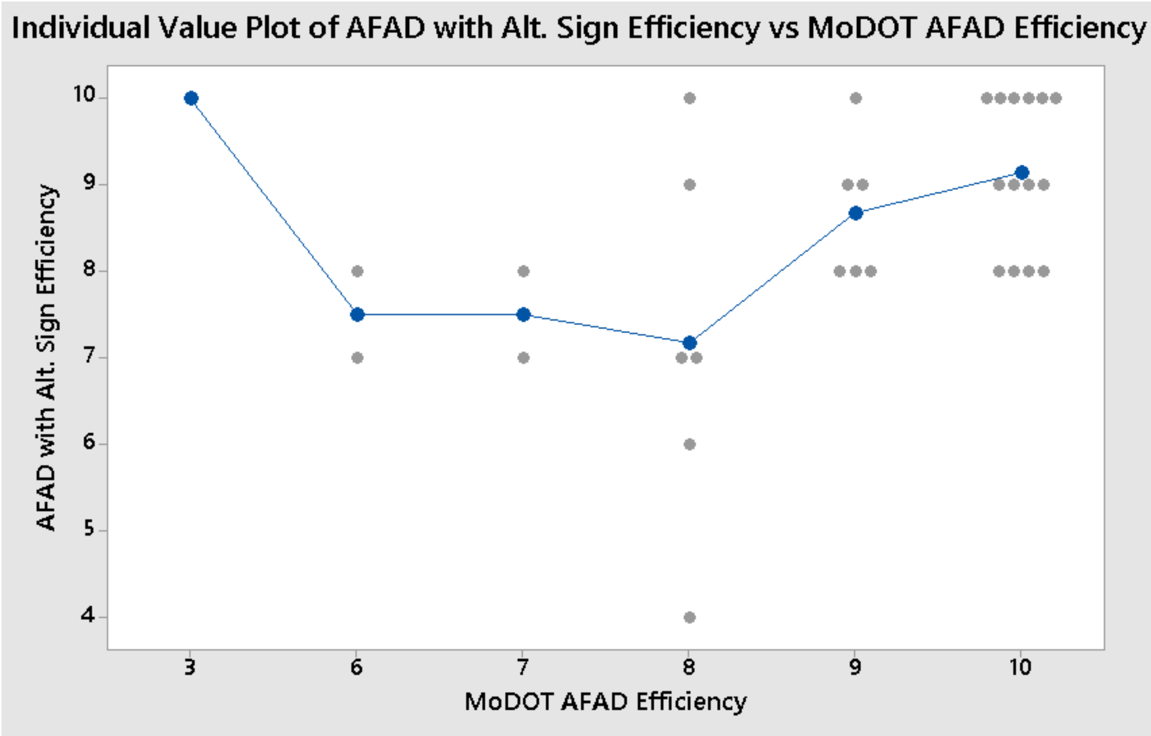


Figure G.4.2 Efficiency: AFAD with Alternative Sign vs. MoDOT AFAD

MoDOT AFAD Efficiency	N	Mean	StDev	95% CI
3	1	10.00	*	(7.50, 12.50)
6	2	7.500	0.707	(5.731, 9.269)
7	2	7.500	0.707	(5.731, 9.269)
8	6	7.167	2.137	(6.145, 8.188)
9	6	8.667	0.816	(7.645, 9.688)
10	14	9.143	0.864	(8.474, 9.811)

Pooled StDev = 1.21459

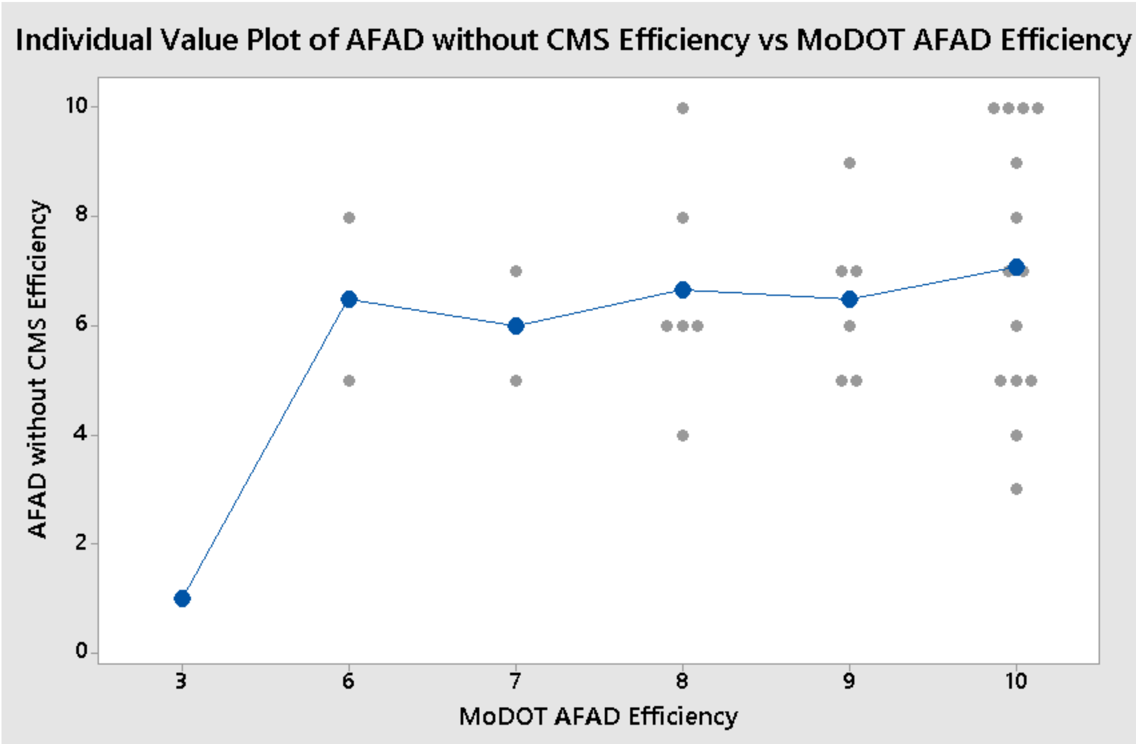


Figure G.4.3 Efficiency: AFAD without CMS vs. MoDOT AFAD

MoDOT AFAD Efficiency	N	Mean	StDev	95% CI
3	1	1.000	*	(-5.112, 7.112)
6	2	7.00	4.24	(2.68, 11.32)
7	2	3.50	2.12	(-0.82, 7.82)
8	6	6.833	2.137	(4.338, 9.329)
9	6	5.17	3.13	(2.67, 7.66)
10	14	5.000	3.113	(3.367, 6.633)

Pooled StDev = 2.96760

APPENDIX H: AFAD SSQ RESULTS IN PERCENTAGE

Table H.0.1 SSQ Results

	General discomfort	Fatigue	Headache	Eye strain
None	78.13%	96.88%	75.00%	71.88%
Slight	18.75%	3.13%	25.00%	25.00%
Moderate	3.13%	0.00%	0.00%	3.13%
Severe	0.00%	0.00%	0.00%	0.00%
	Difficulty focusing	Salivation increasing	Sweating	Nausea
None	84.38%	96.88%	93.75%	81.25%
Slight	12.50%	3.13%	6.25%	12.50%
Moderate	3.13%	0.00%	0.00%	3.13%
Severe	0.00%	0.00%	0.00%	3.13%
	Difficulty concentrating	Fullness of the Head	Blurred vision	Dizziness with eyes open
None	96.88%	87.10%	90.32%	84.38%
Slight	3.13%	9.68%	9.68%	12.50%
Moderate	0.00%	3.23%	0.00%	3.13%
Severe	0.00%	0.00%	0.00%	0.00%
	Dizziness with eye closed	Vertigo	Stomach awareness	Burping
None	93.75%	96.88%	84.38%	93.75%
Slight	6.25%	3.13%	9.38%	6.25%
Moderate	0.00%	0.00%	6.25%	0.00%
Severe	0.00%	0.00%	0.00%	0.00%

APPENDIX I: POST-SIMULATOR SURVEY OF GREEN LIGHT PROJECT

Truck-Mounted Attenuator (TMA) Simulator Survey

Proper communication of work zone information is critical for the safe movement of traffic through work zones. Please tell us your perspective on how to make work zones better.

Daytime TMA	 Figure 1a (green only)	 Figure 1b (green and yellow)	 Figure 1c (yellow and white)	 Figure 1d (green and white)
1.	**Please rate each lightbar color option from [5] Like it very much to [1] Dislike it very much.**			
	[5] [4] [3] [2] [1]	[5] [4] [3] [2] [1]	[5] [4] [3] [2] [1]	[5] [4] [3] [2] [1]
2.	**Please rate all lightbar options from a scale of 10 (highest) to 1 (lowest) with respect to following:**			
Visibility of work zone vehicles				
Awareness of work zones				
Clear recognition of arrow direction				
Easy on the eyes				
Comments				

Nightime TMA	 Figure 2a (green only)	 Figure 2b (green and yellow)	 Figure 2c (yellow and white)	 Figure 2d (green and white)
3.	**Please rate each lightbar color option from [5] Like it very much to [1] Dislike it very much.**			
	[5] [4] [3] [2] [1]	[5] [4] [3] [2] [1]	[5] [4] [3] [2] [1]	[5] [4] [3] [2] [1]
4.	**I think the lights are too bright.**			
	[Disagree] [Neutral] [Agree]	[Disagree] [Neutral] [Agree]	[Disagree] [Neutral] [Agree]	[Disagree] [Neutral] [Agree]
5.	**Please rate all lightbar options from a scale of 10 (highest) to 1 (lowest) with respect to following:**			
Visibility of work zone vehicles				
Awareness of work zones				
Clear recognition of arrow direction				
Easy on the eyes				
Comments				

Please answer questions about your simulator experience.

6. I felt like I was actually there on the highway.
[ ] Strongly Agree [ ] Agree [ ] Neutral [ ] Disagree [ ] Strongly Disagree

7. I felt like I could drive around freely.
[ ] Strongly Agree [ ] Agree [ ] Neutral [ ] Disagree [ ] Strongly Disagree

Please answer the demographic questions below.

8. Age range
[ ] 16-25 [ ] 26-40 [ ] 41-55 [ ] 56-70 [ ] 71-95

9. Gender
[ ] Male [ ] Female

10. My Residency
[ ] Urban [ ] Rural

11. My Regular Vehicle Type
[ ] Passenger Car [ ] Vehicle towing trailer [ ] Delivery/Moving Truck

[ ] Tractor trailer truck [ ] Bus

12. Please enter any additional comments you may have regarding this study.

____

____

____

Please contact Mr. Henry Brown (brownhen@missouri.edu) for additional comments, concerns or information on this survey.

Thank you for completing this survey! We greatly appreciate your time!

APPENDIX J: ADDITIONAL COMMENTS FROM POST-SIMULATOR SURVEY OF GREEN LIGHT PROJECT

Participants commented on both daytime and nighttime configurations. As shown in Table J.0.1 (daytime) and Table J.0.2 (nighttime). The major comments concerned brightness, visibility of the work zone, disability glare, and clarity and effectiveness of the work zone.

Table J.0.1 Daytime comments

	Number of comments					
	Too bright	Too dim	Glare	Distracting	Clear/effective	Blurry/strain on eyes
Amber/white	6	-	2	-	-	-
Green only	-	2	1	-	2	1
Green/amber	1	-	1	-	2	-
Green/white	1	-	1	1	-	1

Table J.0.2 Nighttime comments

	Number of comments			
	Too bright	Hard to see	Glare	Clear/Effective
Amber/white	4	-	2	1
Green only	1	1	-	3
Green/amber	2	1	-	3
Green/white	2	-	2	2

Other overall comments are listed as:

- Green/amber- best mix of visibility and comfort (4).
- Preferred green/amber, but was more used to amber/white (1).
- Green is easy on the eyes at night, but may confuse drivers (2).
- Looking at the arrow more than the lights- only made a difference at night (1)
- Work zone was easy to see, but difficult to tell merge direction (1).
- Most helpful thing was the light bar pulsing in the direction of the arrow with the yellow color (1).

APPENDIX K: AUTONOMOUS TRUCK PLATOON IN WORK ZONES POST-SIMULATOR SURVEY

Figure 1

1. **What is the meaning of figure 1?**

a) Trucks ask us to follow them
b) Trucks are moving together as a team
c) The sign/message makes no sense

2. **Please rate the effectiveness of the signage/message shown in Figure 1.**

[ ] Very Effective [ ] Effective [ ] Neutral [ ] Ineffective [ ] Very Ineffective

Figure 2

3. **What is the meaning of figure 2?**

a) Trucks ask us to follow them
b) Two/four trucks are moving together as a team
c) The sign/message makes no sense

4. **Please rate the effectiveness of the signage/message shown in Figure 2.**

[ ] Very Effective [ ] Effective [ ] Neutral [ ] Ineffective [ ] Very Ineffective

Please read the paragraph and answer questions

A "platoon" means that a team of vehicles are travelling together, and they interact with each other within the platoon. A truck platoon means these trucks are moving together as a team. The display on the back of the trucks indicates either trucks are in a platoon, or the number of trucks in this platoon.

5. **The paragraph was helpful to understand sign/message displayed on truck.**

[ ] Strongly agree [ ] Agree [ ] Neutral [ ] Disagree [ ] Strongly disagree

6. **After education, I feel clearer how to react with the truck platoon.**

[ ] Strongly agree [ ] Agree [ ] Neutral [ ] Disagree [ ] Strongly disagree

7. **Which driver behavior do you think is the safest reaction to truck platoons in/near work zones?**
   a) Slow down and follow
   b) Speed up and bypass before entering work zone
   c) Squeeze in and drive between trucks
   d) I do not know

8. **What is your preference of education?**

[ ] I get educated about trucks moving together as a platoon [ ] No education

9. **What is your preference of display?**

[ ] Display "truck platoon" [ ] Display number of trucks in the platoon [ ] No display

10. **What is your preference as a vehicle driver encountering the truck platoon?**

[ ] Less trucks in the platoon
[ ] More trucks in the platoon

11. **Age range**

[ ] 18-25 [ ] 26-40 [ ] 41-55 [ ] 56-70 [ ] 71-95

12. **Gender**

[ ] Male [ ] Female

13. **My Residency**

[ ] Urban [ ] Rural

14. **My Regular Vehicle Type**

[ ] Passenger Car [ ] Vehicle towing trailer [ ] Delivery/Moving Truck
[ ] Tractor trailer truck [ ] Bus

15. **Please enter any additional comments you may have regarding this study.**

____

APPENDIX L: SAS CODE FOR MIXED MODELS

Model 1:

```
proc mixed method=type3;
      class Configuration Time ID Gender Age;
      model distance = configuration Time;
      lsmeans configuration/adjust=tukey;
      lsmeans time/adjust=tukey;
      run;
```

Model 2:

```
proc mixed method=type3;
      class Configuration Time ID Gender Age;
      model distance = configuration Time;
      random ID;
      lsmeans configuration/adjust=tukey;
      lsmeans time/adjust=tukey;
      run;
```

Model 3:

```
proc mixed method=type3;
      class Configuration Time ID Gender Age;
      model distance = configuration Time Gender AgeGroup;
      lsmeans configuration/adjust=tukey;
      lsmeans time/adjust=tukey;
      lsmeans Gender/adjust=tukey;
      lsmeans AgeGroup/adjust=tukey;
      run;
```

Model 4:

```
proc mixed method=type3;
      class Configuration Time ID Gender Age;
      model distance = configuration | Time | Gender | AgeGroup;
      lsmeans configuration/adjust=tukey;
      lsmeans time/adjust=tukey;
      lsmeans Gender/adjust=tukey;
      lsmeans AgeGroup/adjust=tukey;
      run;
```

Modeling Full Stop Distance (ft.) of AFAD Simulator Study:

```
proc mixed method=type3;
      class Device Understanding Gender Age;
      model Stop = Device Understanding Gender Age Clarity Visibility Safety
Efficiency CMSNecessary OnHighway DriveFreely;
      lsmeans device/adjust=tukey;
      lsmeans Understanding/adjust=tukey;
      lsmeans Gender/adjust=tukey;
      lsmeans Age/adjust=tukey;
      run;
```

VITA

Siyang Zhang received her Ph.D. degree in Transportation Engineering, with a designated Ph.D. minor in Statistics, in University of Missouri-Columbia in May 2020. She obtained her master degree in Geography with emphasis in Geospatial Information Science (GIS) in December 2016. She gained double bachelor degrees in Electrical Engineering from University of Missouri-Columbia and North China Electric Power University in May 2013 and July 2013, respectively, with Cum Laude award.

In her doctoral study, Siyang Zhang focused on investigation of smart work zone technologies, aiming to improve work zone safety. She designed, developed, and conducted simulator studies and field studies, along with surveys. Her original and significant contributions included collaboration of simulator and field studies, and building statistical models to incorporate simulator data, post-simulator survey response, field data, and field survey response together, fitting quantitative and qualitative data into one comprehensive model.