U.S. Department of Transportation

Federal Railroad Administration

Driver Behavior at Highway-Rail Grade Crossings Using NDS and Driving Simulators

Office of Research,
Development,
and Technology
Washington, DC 20590

DOT/FRA/ORD-20/47

Final Report
December 2020

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	December 2020	Technical Report, Sept. 2016–Oct. 2018

4. TITLE AND SUBTITLE

Driver Behavior at Highway-Rail Grade Crossings Using NDS and Driving Simulators

5. FUNDING NUMBERS

6. AUTHOR(S)

Pasi Lautala, Myounghoon Jeon, David Nelson, Steven Landry, Aaron Dean

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Michigan Technological University
1400 Townsend Drive
Houghton, MI 49931-1200

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Department of Transportation
Federal Railroad Administration
Office of Railroad Policy and Development
Office of Research, Development and Technology
Washington, DC 20590

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

DOT/FRA/ORD-20/47

11. SUPPLEMENTARY NOTES

COR: Starr Kidda

12a. DISTRIBUTION/AVAILABILITY STATEMENT

This document is available to the public through the FRA eLibrary.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

Researchers at Michigan Technological University used Second Strategic Highway Research Program (SHRP2) naturalistic driving study (NDS) data and a driving simulator to perform a quantitative evaluation of driver behavior at highway-rail grade crossings (HRGCs). They developed a 3-point scale to generate driver behavior scores for over 9,000 NDS traversals. They used the mean scores to perform statistical comparisons of driver behavior at HRGCs with different traffic control devices (TCDs), with and without an accident history, and in various environmental conditions. They also simulated two HRGCs in a driving simulator and compared the driver behavior scores between naturalistic and simulated environments.

This study revealed that most drivers did not visually scan for trains and did not prepare to stop, regardless the type of warning device present at the crossing or the environmental conditions at the time of traversal. The results were fairly consistent in both NDS and simulated approaches. The NDS data analysis showed little statistical difference in driving behavior between any of the TCDs analyzed. The only exceptions were the significantly higher mean scores at passive HRGCs equipped with stop signs and the higher mean scores for day versus night traversals.

14. SUBJECT TERMS

Grade crossing, traffic control device, average annual daily traffic, Federal Highway Administration, Federal Railroad Administration, vehicle miles traveled, freight train miles

15. NUMBER OF PAGES

56

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18
298-102

METRIC/ENGLISH CONVERSION FACTORS

ENGLISH TO METRIC

LENGTH (APPROXIMATE)

1 inch (in)	=	2.5 centimeters (cm)
1 foot (ft)	=	30 centimeters (cm)
1 yard (yd)	=	0.9 meter (m)
1 mile (mi)	=	1.6 kilometers (km)

AREA (APPROXIMATE)

1 square inch (sq in, in^2)	=	6.5 square centimeters (cm^2)
1 square foot (sq ft, ft^2)	=	0.09 square meter (m^2)
1 square yard (sq yd, yd^2)	=	0.8 square meter (m^2)
1 square mile (sq mi, mi^2)	=	2.6 square kilometers (km^2)
1 acre = 0.4 hectare (he)	=	4,000 square meters (m^2)

MASS - WEIGHT (APPROXIMATE)

1 ounce (oz)	=	28 grams (gm)
1 pound (lb)	=	0.45 kilogram (kg)
1 short ton = 2,000 pounds (lb)	=	0.9 tonne (t)

VOLUME (APPROXIMATE)

1 teaspoon (tsp)	=	5 milliliters (ml)
1 tablespoon (tbsp)	=	15 milliliters (ml)
1 fluid ounce (fl oz)	=	30 milliliters (ml)
1 cup (c)	=	0.24 liter (l)
1 pint (pt)	=	0.47 liter (l)
1 quart (qt)	=	0.96 liter (l)
1 gallon (gal)	=	3.8 liters (l)
1 cubic foot (cu ft, ft^3)	=	0.03 cubic meter (m^3)
1 cubic yard (cu yd, yd^3)	=	0.76 cubic meter (m^3)

TEMPERATURE (EXACT)

[(x-32)(5/9)] □F = y □C

METRIC TO ENGLISH

LENGTH (APPROXIMATE)

1 millimeter (mm)	=	0.04 inch (in)
1 centimeter (cm)	=	0.4 inch (in)
1 meter (m)	=	3.3 feet (ft)
1 meter (m)	=	1.1 yards (yd)
1 kilometer (km)	=	0.6 mile (mi)

AREA (APPROXIMATE)

1 square centimeter (cm^2)	=	0.16 square inch (sq in, in^2)
1 square meter (m^2)	=	1.2 square yards (sq yd, yd^2)
1 square kilometer (km^2)	=	0.4 square mile (sq mi, mi^2)
10,000 square meters (m^2)	=	1 hectare (ha) = 2.5 acres

MASS - WEIGHT (APPROXIMATE)

1 gram (gm)	=	0.036 ounce (oz)
1 kilogram (kg)	=	2.2 pounds (lb)
1 tonne (t)	=	1,000 kilograms (kg)
	=	1.1 short tons

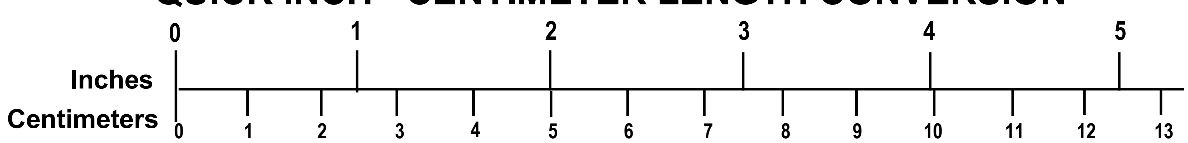
VOLUME (APPROXIMATE)

1 milliliter (ml)	=	0.03 fluid ounce (fl oz)
1 liter (l)	=	2.1 pints (pt)
1 liter (l)	=	1.06 quarts (qt)
1 liter (l)	=	0.26 gallon (gal)
1 cubic meter (m^3)	=	36 cubic feet (cu ft, ft^3)
1 cubic meter (m^3)	=	1.3 cubic yards (cu yd, yd^3)

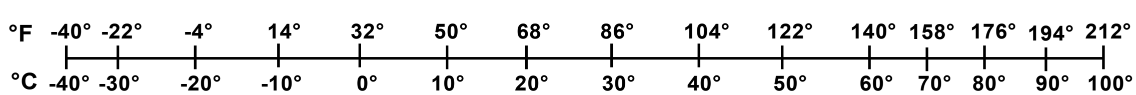
TEMPERATURE (EXACT)

[(9/5) y + 32] □C = x □F

QUICK INCH - CENTIMETER LENGTH CONVERSION



QUICK FAHRENHEIT - CELSIUS TEMPERATURE CONVERSIO



For more exact and or other conversion factors, see NIST Miscellaneous Publication 286, Units of Weights and Measures. Price $2.50 SD Catalog No. C13 10286

Updated 6/17/98

Acknowledgments

The research team would like to acknowledge the following individuals and organizations for their support and contributions to the project:

- The Federal Railroad Administration's Starr Kidda, Chief of the Human Factors Division of the Office of Research, Development and Technology; and Debra Chappell, Transportation Analyst in the Office of Railroad Safety, for technical support
- Virginia Tech Transportation Institute's Miguel Perez, Edie Sears, and the rest of the team for support with the Second Strategic Highway Research Program's naturalistic driving study data
- The National Advanced Driving Simulator team at Iowa State University
- Michigan Tech Transportation Institute researchers (in addition to the contributing authors): Kyle Dick, Modeste Muhire, Darian Reed, Alawudin Salim, and Yugang Wang
- National University Rail Center, the U.S. Department of Transportation's Tier 1 University Transportation Center

Contents

Illustrations

Tables

Executive Summary

Michigan Technological University (Michigan Tech) used the Second Strategic Highway Research Program (SHRP2) naturalistic driving study (NDS) database and a driving simulator to perform a quantitative evaluation of driver behavior at highway-rail grade crossings (HRGCs). The evaluation took place from September 2016 through October 2018. The Federal Railroad Administration sponsored this study to quantify the level of defensive driving behavior during HRGC traversals. The researchers developed a 3-point driver behavior score and automatic data processing application to generate scores for over 9,000 traversals. The mean scores were used to perform statistical comparisons of driver behavior at HRGCs with different traffic control devices (TCDs) and between HRGCs with and without past accidents. The research team also explored whether any trending could be identified among parameters categorized as critical to safety in previous studies. Finally, the team simulated two HRGCs in a driving simulator and compared the driver behavior scores between a naturalistic and artificial environment.

The investigation of over 9,000 NDS traversals revealed that most drivers did not visually scan for trains and did not prepare to stop, regardless the type of warning device present at the crossing or the environmental conditions at the time of traversal. The results were moderately consistent in both NDS and simulated approaches.

The main findings of the study include:

- The NDS data analysis showed little statistical difference in driving behavior between any of the TCDs analyzed. The only exceptions were the significantly higher mean scores at passive HRGCs equipped with stop signs and consistently higher mean scores for day versus night traversals.
- Researchers found no significant evidence of systematic driver behavior differences when comparing different demographics, such as behavior between genders or between age groups.
- Trending analysis based on average annual daily traffic, trains per day, and train/highway speed impact on driver behavior score provided interesting observations, but statistical analysis were not conducted due to the exploratory nature of the analysis.
- Simulator testing and NDS data analysis provided similar results, but there were more significant differences between scenarios in the NDS than in the simulated setting.

A better understanding of driver behavior at HRGCs would help predict situations when drivers are less cautious and could be at risk of accidents. The SHRP2 NDS database can improve that understanding. However, instead of relying on a simple 3-point scale to evaluate driver behavior, future research should concentrate on analyzing specific parameters, or parameter clusters, that are expected to have the greatest impact on the behavior. Researchers also believe that techniques such as multivariate analysis, machine learning, and artificial intelligence can be harnessed to assist in driving simulator research.

1. Introduction

1.1 Background

Highway-rail grade crossings (HRGCs) are locations where a highway (i.e., road or street, including its associated pathways and sidewalks) crosses one or more railroad tracks at grade. They may also be called railroad crossings (RC) or level crossings (LC). HRGCs may be public or private. Private HRGCs are not maintained by public highway authorities and are not intended to be used by the public. According to the Federal Railroad Administration (FRA), there were a total of 211,631 HRGCs operating in the United States in 2015, and more than 60 percent of them were considered public [1].

Together with trespassing incidents, HRGC accidents – also called collisions or crashes – between roadway vehicles and trains are the greatest source of injuries and fatalities related to rail transportation in North America. A motorist is 40 times more likely to be killed in a vehicle-train accident than in any other type of highway collision [2]. To illustrate the seriousness of the problem, there were 18,289 collisions between 2008–2017, resulting in 2,250 fatalities and over 8,000 injuries [3]. Because of numerous safety efforts, the total number of HRGC accidents has significantly decreased over the last decades. However, since 2009 the number of HRGC accidents has increased slightly, most likely due to the increased rail and road traffic volumes (see Figure 1) [3].

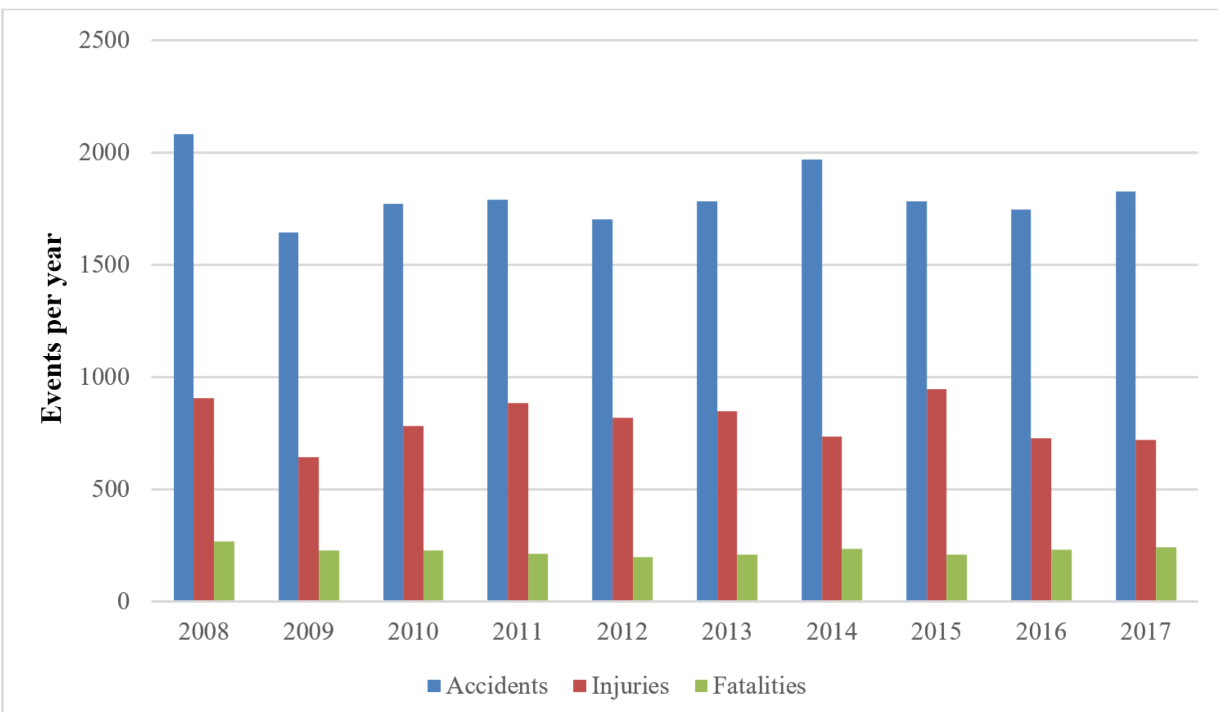


Figure 1. Number of Annual HRGC Injuries and Fatalities, 2008–2017

Key elements in both cause and prevention of accidents is the drivers' behavior and their reactions to the surrounding conditions, and traffic control devices (TCDs) at HRGCs. FRA's 2016 report on HRGC accidents states that 94 percent of train-vehicle collisions can be attributed to driver behavior or poor judgment, implying that risky behavior by drivers, or a lack of defensive driving, is likely to increase the possibility of an accident at HRGCs [4]. Previous studies on HRGC accidents have also indicated several other factors that increase the accident risk at HRGCs. These factors include rail and highway traffic volumes, train speeds, number of tracks and highway lanes, HRGC angle, TCD type, driver demographics, and time of day for the traversal [1].

The long-standing challenge to lower the number of casualties and accidents at HRGCs warrants a consideration of any new methods and technologies to help in the quest toward zero accidents. The research team used two potential approaches, naturalistic driving study data, and driving simulators. Naturalistic driving studies use instrumented vehicles of everyday drivers to quantitatively evaluate the behavior of those drivers.

Driving simulators are used in a variety of research domains to offer insight into driver behavior. They allow research in a controlled environment and provide complimentary technology for naturalistic studies. However, it is important to establish the validity of simulator data as a surrogate measure of real-world behavior in a specific given context before extrapolating the results to inform public policy, or the design of new technology. The researchers used the correlation between naturalistic driving study results and simulated data as an example of such validation process.

1.2 Objectives and Scope

Previous research on HRGC safety has often used accident reports to predict situations when HRGC accidents are more likely to happen, or used the traffic volumes and infrastructure conditions as an indicator of risk at HRGCs. In other words, many past studies on HRGC safety, especially those considering the role of human behavior, have concentrated on after-the-fact analyses of accidents [5] [6]. Other methods, such as external video recordings and roadside or in-vehicle observations have also been used, but those efforts have often provided only partial data of the driving event (i.e., internal or external) and tend to have limited sample sizes for developing large-scale trends.

A few past studies have evaluated naturalistic driving data. These have examined motorist behavior by installing video cameras and sensors in automobiles and analyzing the drivers' actions. For example, FRA conducted an evaluation of driver behavior at HRGCs in a 2010 study involving light vehicle drivers. The data included information about drivers' activities, driver and vehicle performance, driving environment, and vehicle location at or on approach to HRGCs. [7].

This two-phase project continues HRGC safety research that uses direct and detailed observation of the drivers. The overall objective is to investigate driver behavior at HRGCs using two distinct but complimentary techniques. Phase 1 takes advantage of the extensive Second Strategic Highway Research Program (SHRP2) naturalistic driving study (NDS) database [8]. The NDS study approach allows for systematic analysis of in-vehicle video and other sensors for the direct observation of drivers during typical driving activities at HRGCs. Researchers used the data, together with an evaluation methodology developed at Michigan Technological University (Michigan Tech), as part of the project in an attempt to quantify the level of defensive driving behavior during HRGC traversals. More specifically, the analysis concentrated on:

- Driver response to different TCDs at HRGCs
- Comparison of driver behavior at HRGCs with and without accidents between 2000 and 2010
- Exploration of the use of NDS data for trending analysis

In Phase 2, researchers created simulated scenarios that resembled environments found in the NDS data. Driver behavior data from these simulated scenarios was collected and compared with the NDS datasets. More specifically, the objectives were to:

- Select two HRGCs from the NDS dataset and recreate them in a simulated setting.
- Recruit student drivers to participate in a driving simulation study where drivers are exposed to different HRGCs.
- Compare and contrast driver behavior between the NDS and simulator datasets, using the previously developed driver behavior score.

1.3 Organization of the Report

This report first summarizes and discusses the research data, activities and outcomes for Phase 1 in Section 2 through 4, followed by a similar summary for Phase 2 in Section 5. Section 6 explains the study limitations, while Section 7 offers a conclusion and discussion of potential future research.

2. Phase 1: Data Sources and Methodology

Figure 2 outlines the data sources and process flow for Phase 1 research. Each process component is explained in more detail in the following sections.

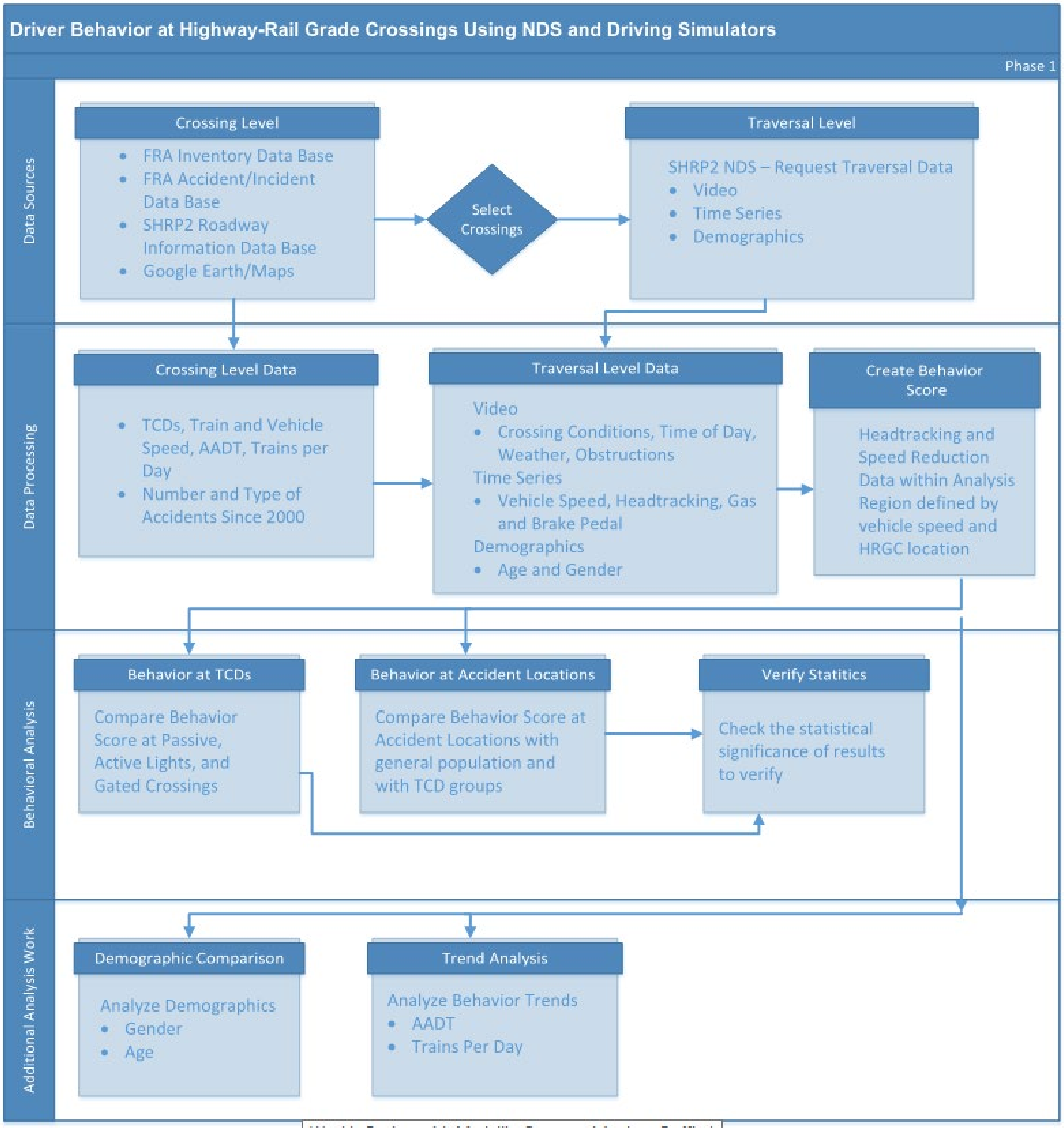


Figure 2. Phase 1 Data Sources and Research Process

2.1 Data Sources

The study used four main data sources: the SHRP2 NDS [8], SHRP2 Roadway Information Database (RID) [9], the FRA grade crossing inventory database [10], and the FRA accident/incident database [11]. Google Maps, Google Earth, and forward video streams of the NDS data were used to verify TCDs at selected HRGCs.

2.1.1 SHRP2 NDS and SHRP2 RID Databases

The NDS, funded through the Transportation Research Board under the National Academies of Science, captured the unsupervised driving performance of participants in Florida, Indiana, New York, North Carolina, Pennsylvania, and Washington. The study was conducted in 2010–2013 and included more than 5 million trips by approximately 3,500 participants [8]. The data from the NDS trip database used for this study included detailed sensor information on vehicle location, brake and throttle position, vehicle speed and acceleration, and driver demographics (Table 1). It also included front and rear video feeds from the NDS video files, and head rotation and position derived from a face video feed (Table 2). The database is stored in a secure data enclave at the Virginia Tech Transportation Institute and is accessible across the U.S. through a data use license.

Table 1. Time Series Data from NDS Trip Database

Head Rotation and Position	Vehicle Speed	Vehicle Accelerations
Global Positioning System (GPS) Location	Day and Time	Brake and Throttle
Age Group	Gender	Forward and Rear Video

Table 2. Data Collected from NDS Video Files

Weather	Day/Night Conditions	Traffic Conditions
Crossing Position	Other Crossing Features	Crossing Conditions
Sun in Face		

The RID was also developed under the SHRP2 to provide roadway and route information on trips taken by the NDS participants [9]. Michigan Tech used the RID to identify 1,017 public HRGCs traversed by the NDS study participants (Table 3) and selected specific HRGCs from the sample for the analysis in this project.

Table 3. Number of HRGCs per State in the RID

State	HRGCs in NDS
Florida	295
Indiana	104
New York	181
North Carolina	168
Pennsylvania	61
Washington	208
Total	**1,017**

2.1.2 FRA Grade Crossing Inventory Database and Accident/Incident Database

Each of the over 200,000 HRGCs in the FRA inventory database is identified by its FRA crossing ID and described by different data fields that provide HRGC information ranging from ownership to field configuration. The researchers used the crossing ID as a linking field in a programming algorithm to match the available data in the FRA inventory with the corresponding HRGCs in the NDS database. The absence/presence of active warning devices (i.e., lights and/or lights and gates) were used to categorize HRGCs based on their TCDs.

The FRA accident/incident database includes information about reported accidents at HRGCs. This source includes several decades of historical data and was used to identify HRGCs with both an accident history between 2000 and 2010 and traversal data in the NDS. The pertinent data fields used in the study from the two FRA databases are presented in Table 4 and Table 5.

Table 4. Data Collected from the FRA Crossing Inventory Database

Crossing ID	Crossing Type	Crossing Angle
Traffic Control Device	Pavement Marking	AADT
Train Traffic Levels	Latitude and Longitude	Number of Tracks
Number of Roadway Lanes		

Table 5. Data Collected from the FRA Accident/Incident Database

Number of Accidents	Date of Accident	Type of Accident

2.1.3 Google Maps and Forward Cameras and Selection of HRGCs for Analysis

The research team used Google Maps and Google Earth to perform initial verification of TCD status at all 1,017 HRGCs during the NDS study period. The crossing information was later confirmed using the forward-facing video segments of traversals.

2.2 Data Acquisition and Processing

To select a diverse group of HRGCs for the analysis, the researchers used information from the RID and the FRA crossing inventory database. Several key parameters included in the selection process were the type of TCDs, configuration of nearby intersections, and the number of accidents in recent years. Since there was a limited number of passive HRGCs with traversals available in the NDS, all passive HRGCs in all six States were selected for the analysis.

Researchers then limited the study area to New York, Indiana, and Florida; they then selected all 55 HRGCs with flashing lights but no gates for the analysis. For HRGCs with lights and gates, there were over 400 potential candidates – thus, the researchers used proximity to roadway intersections, HRGC angle, and number of accidents at the HRGC as the main selection criteria.

The final list totaled 306 HRGCs – 199 with lights and gates, 55 with flashing lights only, and 54 with passive warning devices (Figure 3).

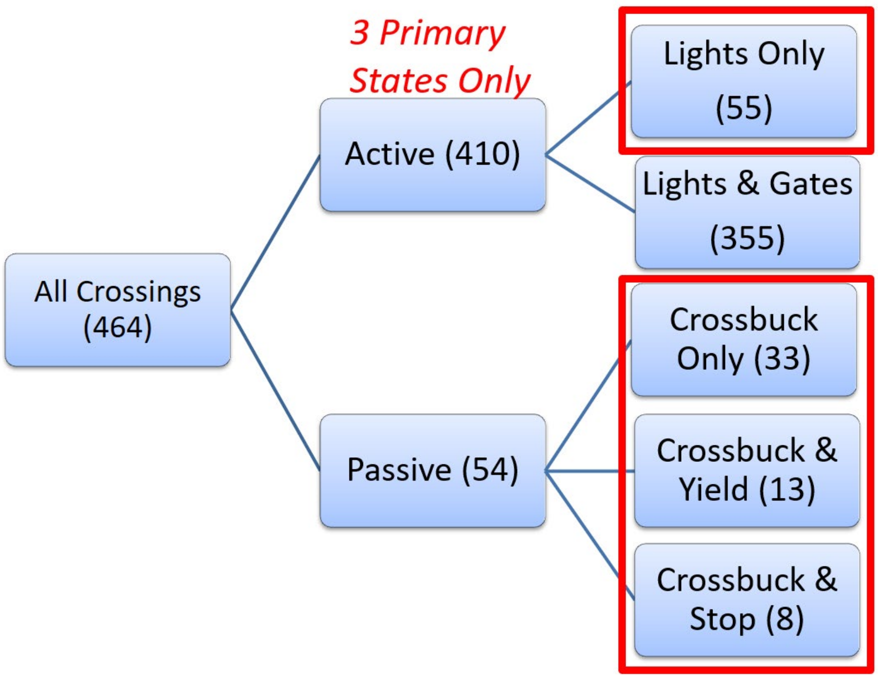


Figure 3. HRGC Selection Process

The requested dataset from the NDS database included an average sample of approximately 40 individual traversals per HRGC. This calculation provided statistically valid results, as described in more detail in a paper by Muhire et al. [13]

Once the HRGCs and sample sizes were determined, the final data request from the NDS data archive included data from almost 13,000 individual traversals. The research team submitted the data request in three batches, which allowed them to begin analysis work earlier and modify the later requests to achieve a more representative sample. The team linked the traversal data to the specific HRGC and stored the compiled dataset in an integrated database for further analysis.

2.2.1 Developing the Behavior Score

Previous research at Michigan Tech developed a methodology to quantitatively evaluate driver behavior when approaching a HRGC [14]. Although the evaluation originally used the term "compliance score," it was later changed to "behavior score," a term considered more appropriate for the analysis.

The behavior score translated qualitative driver behavior data into a quantitative 3-point behavior score suitable for statistical analysis by quantifying two types of actions (see Table 6). A higher behavior score was considered an indication of more defensive driving behaviors.

Table 6. Driver Behavior Score Calculation

Driver Behavior Action	Points Awarded
Visually scan for train to the right (> 8 degrees)	+1
Visually scan for train to the left (> 8 degrees)	+1
Identifiable speed adjustment (> 10% reduction)	+1
Total Possible Score	**+3**

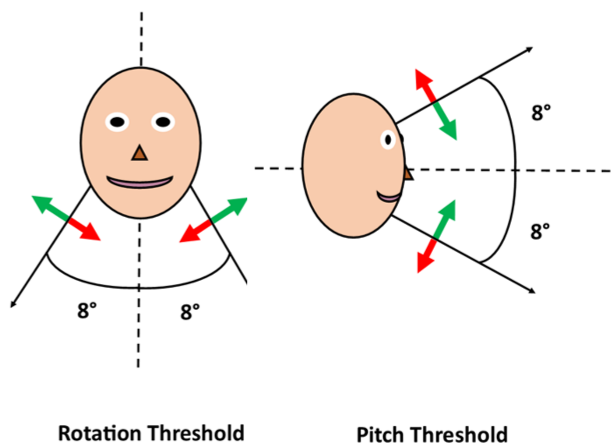


Figure 4. Rotation and Pitch Thresholds Used in Head Tracking

Researchers used head rotation and position to evaluate visual train scanning behavior. Since a train can come from either direction at an HRGC, one point was awarded for scanning in each direction. Scanning scores were based on the headtracking data from the NDS that provides horizontal and vertical head rotation and position data. They used smoothing and interpolation techniques to fill in areas where the original NDS data had gaps [15]. A scan was considered successful if the driver's head rotation exceeded 8 degrees from the baseline after the smoothing—typically the roadway center—in either horizontal direction. To ensure the driver was not looking inside the car, the scan was only awarded points if a pitch, or vertical scan, did not exceed 8 degrees from the baseline (Figure 4). More details of the scanning score process can be found in papers by Lautala et al. [16] [17].

The second component of the behavior score was based on an identifiable speed adjustment that indicated a driver's recognition of approaching a HRGC and preparation to stop, as necessary. The researchers tested three different methods to determine the speed adjustment. The early analysis identified instances where the speed profile showed a deceleration of 2 ft/s^2 within a specific approach zone before the HRGC. Pedal movements were also investigated as an indicator, with the score based on removing the foot from the gas pedal and applying the brake. This proved problematic, as such data was not consistently available from the NDS data.

The research team finally settled on using a third alternative for the analysis, defined as a minimum 10 percent speed reduction within the analysis zone. The analysis zone was determined for each traversal based on the stopping distance required at the entry speed (i.e., the drivers' speed as they approached the HRGC), and the reaction time obtained from the Manual for Uniform Traffic Control Devices (MUTCD), as shown in Figure 5 [18]. The team first identified the last possible braking point for a driver to safely come to a complete stop before reaching the HRGC. The braking point was based on a vehicle deceleration of 11 ft/s^2 per the MUTCD and also became the end point for the analysis range. The researchers allowed drivers 5 seconds before the braking point to complete the proper actions (i.e., head rotation and speed adjustment), demonstrating their preparation for the HRGC traversal. The MUTCD suggests using a reaction time of 2.5 seconds from when the driver can see an obstruction until the driver takes an action in response to the obstruction. The team provided drivers with twice that time period to both recognize the potential danger and to react with necessary precautions.

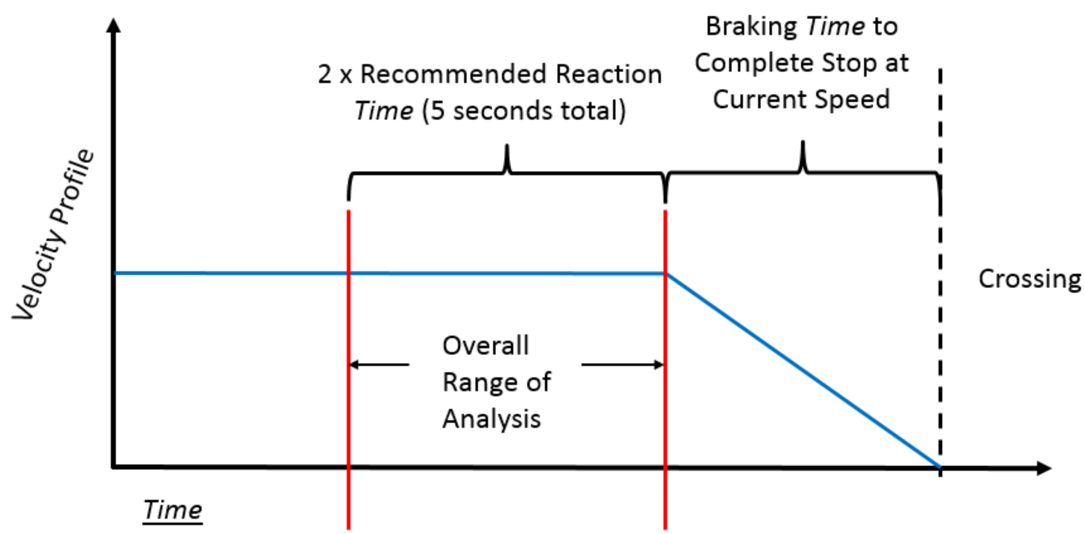


Figure 5. Driver Behavior Score Analysis Zone

Figure 6 and Figure 7 provide examples demonstrating the calculation of the behavior score. In Figure 6 the top graph shows a decrease in velocity of nearly 10 mph (>20 percent), and the middle graph shows a scan to the right (approximately 9 seconds before arrival to the HRGC), and a scan to the left (approximately 7 seconds before). The bottom graph shows that the vertical head position stays within the 8-degree range during both horizontal scans. In this case, the driver received a full 3 points for the traversal.

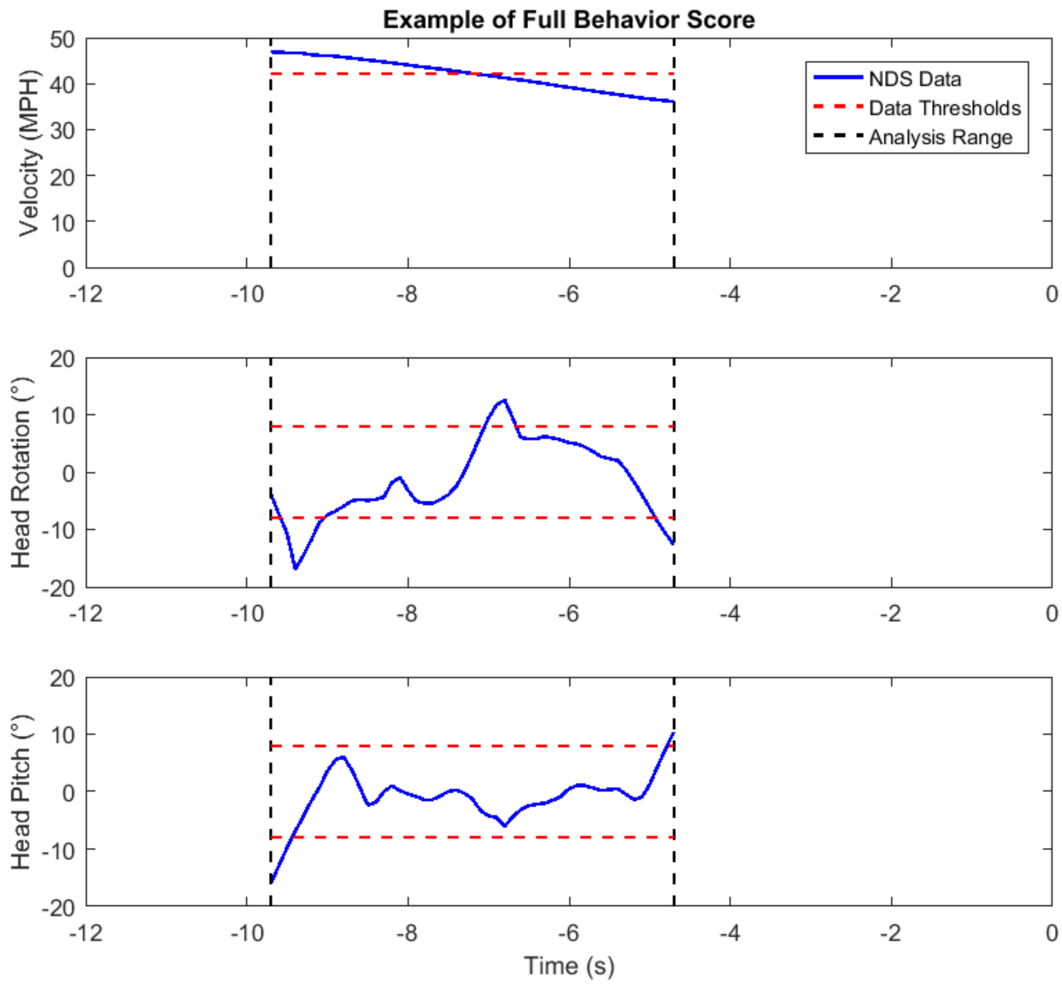


Figure 6. Full (3-Point) Behavior Score Example

Figure 7 shows little change in speed during the analysis period, and the horizontal head rotation stays inside the 8-degree window, suggesting no visual scanning. The pitch would be acceptable, but this traversal would be consistent with a driver looking forward at the road ahead during the entire analysis range. The behavior score for this traversal is zero.

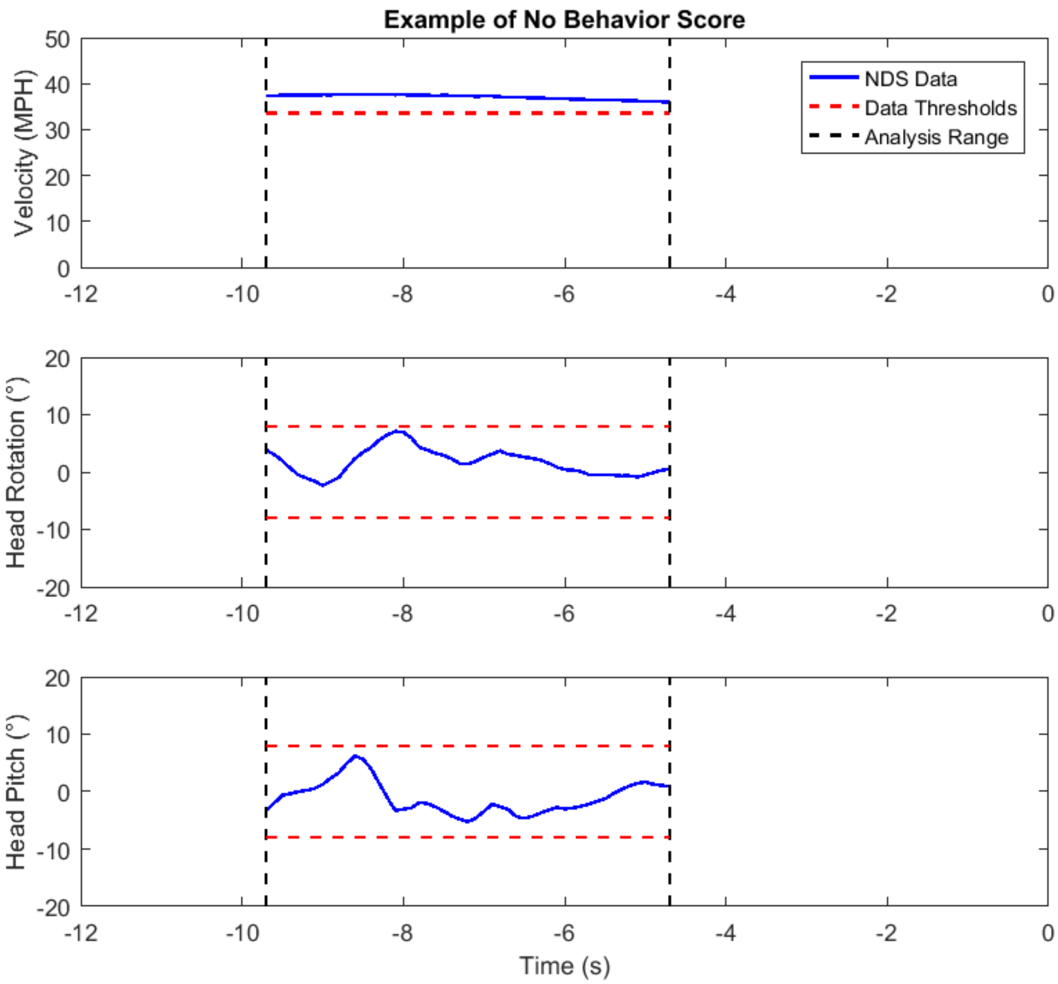


Figure 7. Zero Point Behavior Score Example

2.3 Statistical Analysis

The main objective of the study was to compare driver behavior scores at locations with different TCDs. The researchers also compared behavior between accident and non-accident locations using the mean driver behavior scores at each HRGC as the primary value for comparisons, and conducted a series of paired Welch's t-tests to identify whether the means were statistically different.

The use of results from previous small-scale studies aided in establishing a target sample size of 60 HRGCs for each type of TCD in the analysis. This sample size was determined sufficient to verify the statistical validity of the results at the 90 percent confidence level.

Note that all crossings with the same type of TCD were considered "equal" in this study. This means that other potential differences, such as urban/rural location, number of highway lanes, highway speeds, etc., were not used to categorize the HRCSs in smaller subgroups for the TCD and accident analysis. Although it reduced the homogeneity of the HRGCs in each category, it also ensured the largest possible sample sizes.

3. Results and Data Analysis

The researchers reviewed over 12,000 traversals from 306 crossings in the analysis. However, the records from numerous traversals proved unusable due to incomplete data, such as the missing GPS location of the vehicle. After data reduction to remove such records, 9,128 traversals from 286 HRGCs were determined to have a sufficient level of data to reliably conduct the analysis. The results presented in this section, as well as the statistical test results, are based on this final set of traversals.

Results include the mean behavior scores calculated for the TCD groupings, including separate calculations for the accident locations. The results also provide a breakdown between visual scanning and speed adjustment scores and the sample sizes of each category. This study also presents the results of exploratory trending analysis that uses highway and train speeds and highway and train traffic volumes as main parameters, although this report does not include statistical analysis. A summary of results from a parallel study that investigated the effects of environmental conditions and demographics on driver behavior is also provided.

3.1 Summary of Behavior Scores

Figure 8 shows the mean behavior scores in each TCD category for all crossings and separately for the accident locations. Mean scores between accident and non-accident locations were fairly similar across most categories, except for passive HRGCs with crossbuck and yield signs. However, the accident location dataset for this condition included only one HRGC.

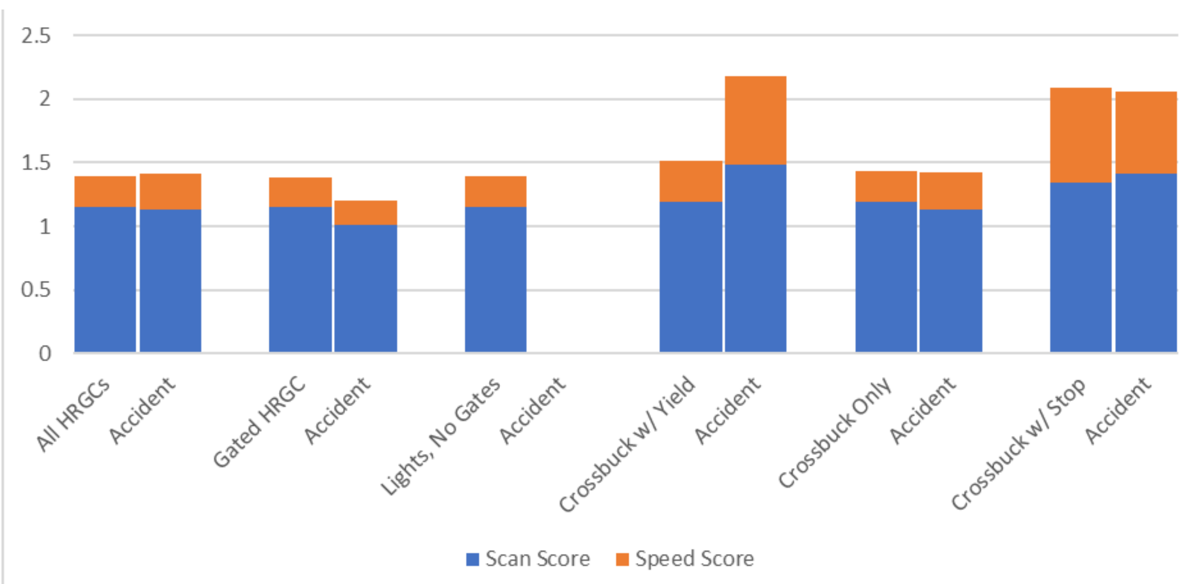


Figure 8. Mean Driver Behavior Scores by TCD for Accident and Non-Accident HRGCs

Table 7 and Table 8 present the mean driver behavior scores, categorized by the prevailing TCDs at the HRGC during the NDS data collection. The mean values and corresponding standard deviations are provided separately for the visual scan and the speed adjustment and combined in the total score. The number of crossings included in the category is in the final column. The results show that the mean speed and visual scan scores were similar within each main category (active vs. passive TCDs), except for the stop conditions.

Table 7. Mean Driver Behavior Scores by TCD

TCD Type	Scan Score	Speed Score	Total Score	# HRGCs
Overall – All HRGCs	1.157	0.238	1.395	286
Std Dev	0.200	0.231	0.340	
Gated HRGC	1.148	0.233	1.381	205
Std Dev	0.193	0.229	0.329	
Lights, No Gates	1.153	0.243	1.395	51
Std Dev	0.202	0.233	0.358	
Crossbuck with Yield	1.189	0.329	1.519	7
Std Dev	0.669	0.366	0.818	
Crossbuck Only	1.191	0.239	1.429	23
Std Dev	0.176	0.248	0.355	
Crossbuck with Stop	1.348	0.745	2.09	5
Std Dev	0.247	0.142	0.16911	

Table 8 shows the behavior scores for the HRGCs reporting accidents between 2000 and 2013. Since the sample size of some categories was small, the value of the statistical analysis was also reduced. HRGCs with only flashing lights and no gates were excluded, as the dataset had no accidents in such locations.

Table 8. Mean Driver Behavior Scores for Accident Locations

TCD Type	Scan Score	Speed Score	Total Score	# HRGCs
Overall – All HRGCs	1.130	0.284	1.403	15
Std Dev	0.704	0.355	0.859	
Gated HRGC	1.013	0.191	1.204	4
Std Dev	0.764	0.329	0.895	
Crossbuck Only	1.133	0.288	1.492	4
Std Dev	0.61	0.15	0.65	
Crossbuck with Yield	1.485	0.697	2.182	1
Crossbuck with Stop	1.414	0.648	2.062	2
Std Dev	0.093	0.101	0.008	

3.2 Summary of Statistical Test Results for TCDs and at Accident Locations

Researchers used a set of paired Welch's t-tests to verify the statistical significance of the behavior score comparisons (Table 9).

Most t-tests failed to reject the null hypothesis, meaning the average values shown were close enough for the means to be equal. The only pairs where the null hypothesis was rejected (i.e., results were statistically different) were comparisons between other TCDs and passive HRGCs with stop signs (bold rows in Table 9).

Table 9. Welch's T-test for Compararison of TCD Conditions

Comparison Pair*	**Scan**	**Score**	**Speed**	**Score**	**Total**	**Score**
	p	**ES**	**p**	**ES**	**p**	**ES**
Gates – Lights	0.7839	0.04	0.8003	0.04	0.8737	0.03
Gates – Passive	0.5942	0.11	0.1694	0.30	0.0961	0.38
Gates – Yield	0.9146	0.04	0.1421	0.83	0.2336	0.92
Gates – Cross	0.9126	0.03	0.5411	0.14	0.2808	0.22
Gates – Overall	0.8122	0.02	0.6705	0.04	0.7805	0.03
Gates – Stop	**0.0000**	**1.10**	**0.0000**	**2.64**	**0.0000**	**2.42**
Lights – Passive	0.7766	0.07	0.3065	0.24	0.1690	0.33
Lights – Yield	0.4633	0.35	0.1658	0.72	0.2353	0.80
Lights – Cross	0.9814	0.02	0.7056	0.10	0.4162	0.20
Lights – Overall	0.8880	0.02	0.9853	0.01	0.9223	0.02
Lights – Stop	**0.0000**	**1.10**	**0.0000**	**3.01**	**0.0000**	**2.53**
Passive – Yield	0.5647	0.26	0.3515	0.45	0.4700	0.42
Passive – Cross	0.7731	0.08	0.5961	0.15	0.5586	0.16
Passive – Overall	0.6622	0.09	0.2241	0.26	0.1313	0.33
Passive – Stop	**0.0082**	**0.69**	**0.0000**	**3.00**	**0.0000**	**2.26**
Yield – Cross	0.4663	0.34	0.2320	0.61	0.3469	0.62
Yield – Overall	0.4242	0.38	0.1578	0.77	0.2415	0.86
Yield – Stop	**0.9099**	**0.09**	**0.0060**	**2.84**	**0.0364**	**1.71**
Cross - Stop	**0.0004**	**0.92**	**0.0000**	**3.21**	**0.0000**	**2.54**
Cross – Overall	0.9853	0.01	0.7054	0.10	0.3720	0.18
Overall – Stop	**0.0000**	**1.02**	**0.0000**	**2.51**	**0.0000**	**2.27**

***** *Definitions: Gates – active HRGCs with gates; Lights – active HRGCs with lights, but no gates; Passive – passive HRGCs, excluding those with stop signs; Yield – passive HRGCs with yield signs; Cross – passive HRGCs with a crossbuck only; Stop – passive HRGCs with a stop sign; Overall – all HRGC locations together; p – probability that both means in the pair are the same; ES – Effect Size***

Similar tests with the accident locations also failed to reject the null hypothesis between TCDs and between accident locations and the general population of HRGCs (Figure 10). There were no accident locations in the NDS dataset for HRGCs with active lights but no gates, and there was only one accident location with a yield sign, so the t-test could not be calculated.

Table 10. Welch's T-test Results for Comparing Accident to Non-Accident Locations

Comparison Pair*	Scan Score		Speed Score		Total Score	
	p	ES	p	ES	p	ES
Overall – All Locations	0.8843	0.11	0.6268	0.19	0.9719	0.02
Gates & Lights	0.7473	0.64	0.8158	0.18	0.7192	0.51
Crossbuck Only	0.8623	0.22	0.6100	0.21	0.8622	0.16
Crossbuck & Stop Sign	0.9300	0.29	0.3849	0.72	0.3694	0.19

See Table 9.

3.3 Trending Parameter Analysis

The research team also performed a preliminary trending analysis on how environmental factors and parameters influence driver behavior scores, concentrating on factors that previous research showed affect the accident risk at HRGCs. To date, these analyses have been exploratory in nature and as such are not considered statistically valid.

The team divided the analysis to two categories: 1) factors that are “crossing-specific” (Category 1), such as AADT, train volumes, and train and highway speeds, and 2) factors that are “traversal-specific” (Category 2), such as weather, time of day, and driver demographics (age and gender).

3.3.1 Category 1 – Crossing Specific Parameters

Category 1 analysis concentrated on investigating whether differences in specific parameters that vary between HRGCs, but remain fixed within a single HRGC, have an effect on mean driver behavior scores. Analysis in Category 1 was based on single factor analysis only, so the exclusion of other factors limits the robustness of the analysis. This could be improved through future multivariate or clustering analysis.

3.3.1.1 *Highway Traffic Levels (AADT)*

Figure 9 shows the average visual scan, speed, and total driver behavior scores versus average annual daily traffic (AADT) data, obtained from the FRA inventory database. Trend lines for both speed and visual scanning scores show a slight decline, as traffic levels increase.

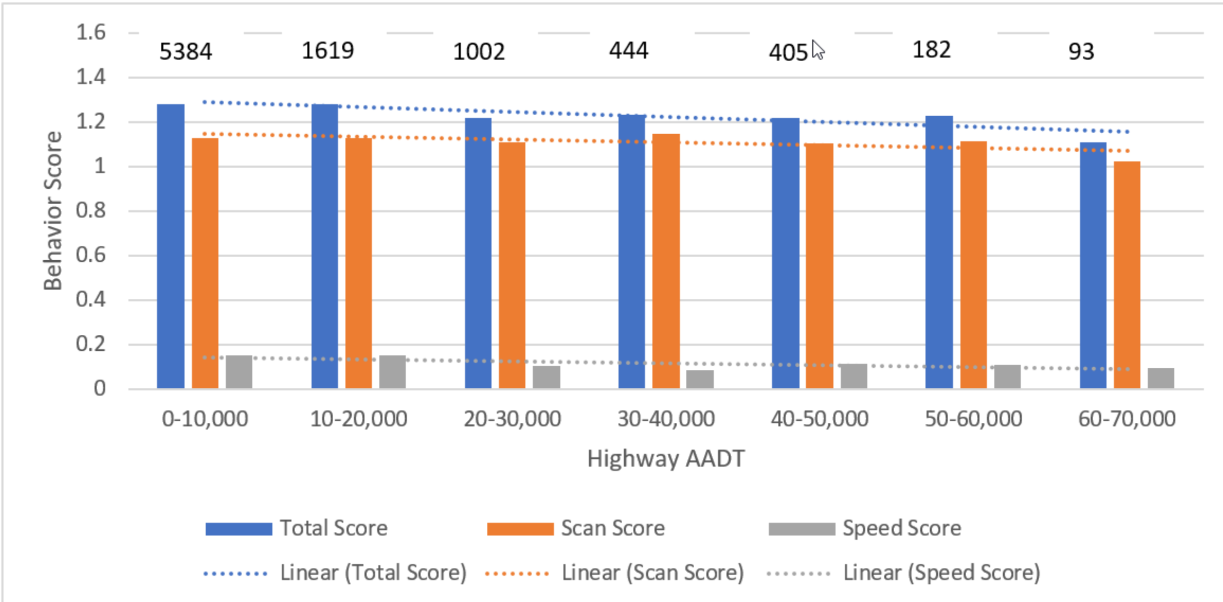


Figure 9. Visual Scan and Speed Scores by AADT (Sample Size per Category on the Top)

3.3.1.2 Train Traffic Volumes

The trains per day (TPD) value is the sum of the day and night through trains from the FRA inventory database. Figure 10 presents the mean driver behavior scores. Note the gap in data between 25 and 50 trains per day, as no HRGCs had TPD values in that range. There were nearly 4,200 traversals at HRGCs with no through trains reported in the FRA inventory, reducing the sample size for the analysis. The analysis reveals increased speeds and visual scan scores as TPD value increased, although the latter occurred at a slower rate.

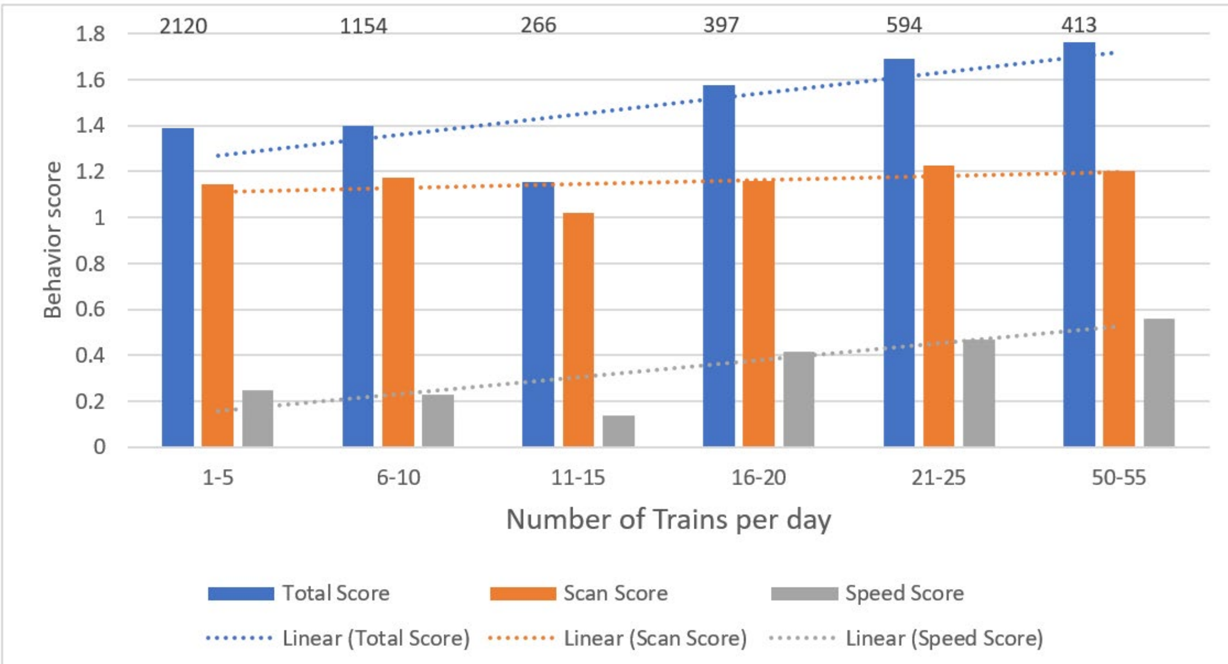


Figure 10. Scan, Speed, and Total Scores by Trains per Day (Sample Size per Category on the Top)

3.3.1.3 Highway Speed Limit

Figure 11 presents the mean behavior scores based on highway speed limit. There was a decrease in the scanning and speed scores as the highway speed increased. However, the data on the low end was limited, as the 10 mph and 15 mph analysis were based on less than 50 traversals each from a single HRGC. More than 1,000 data records were excluded; they did not have a highway speed value posted in the FRA database.

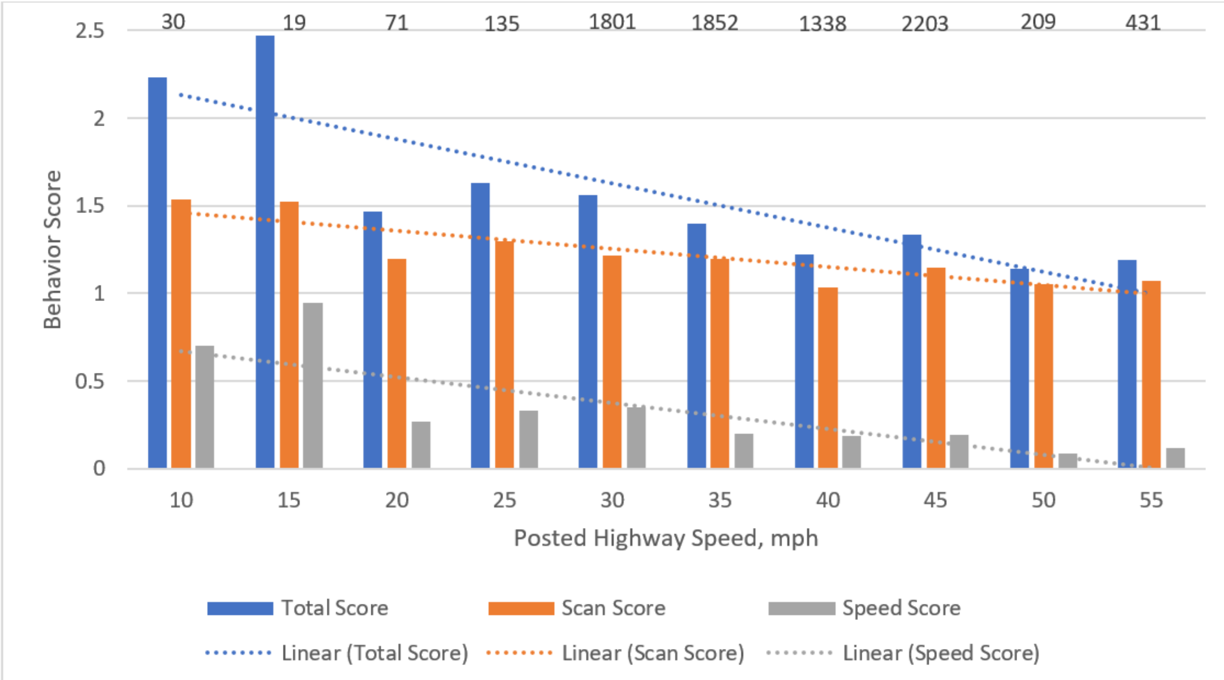


Figure 11. Behavior Scores by Posted Highway Speed (Sample Size per Category on the Top)

3.3.1.4 Train Speed

Figure 12 compares the driver behavior scores with the train speed listed in the FRA inventory. The scanning score appeared to be consistent across the range of train speeds, but the speed score increased with higher train speed. A total of 107 traversals were omitted at these HRGCs. Note the jump in speed from 60 mph to 79 mph, as the dataset had no records with 65 or 70 mph posted speeds.

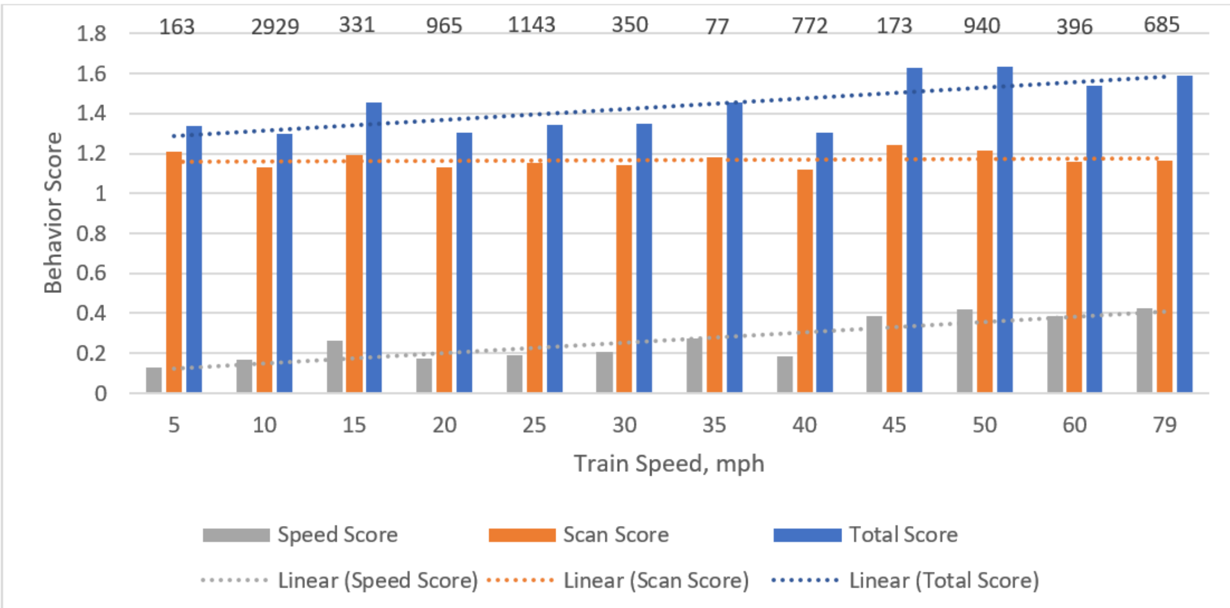


Figure 12. Behavior Scores by Train Speed (Sample Size per Category on the Top)

3.3.2 Category 2 – Traversal-Specific Parameters

Analysis on Category 2 parameters were conducted under a parallel project and concentrated on environmental factors (weather and time of day) and driver demographics (gender and age). These parameters do not remain fixed within a specific HRGC, but may vary between traversals based on prevailing conditions. The following is a summary of the findings from the analysis, including the results of the statistical tests. A more detailed description of the analysis is available in a report by Salim [18].

3.3.2.1 Weather Conditions

Figure 13 shows the sample size and drivers' mean behavior scores based on weather conditions. The data indicated that drivers received the lowest behavior scores in fog – followed by rain, cloudy, clear, and snow conditions. However, no conclusions could be made about drivers' behavior in fog due to the small sample size.

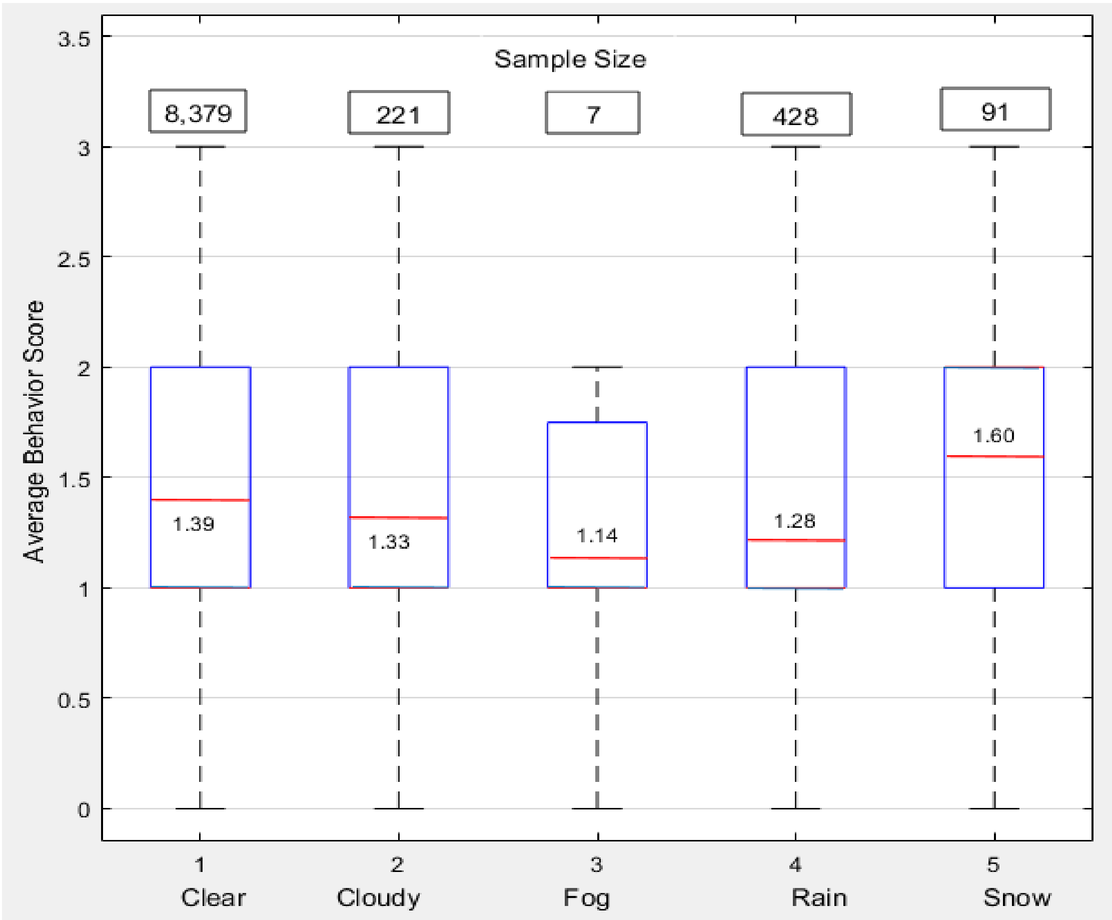


Figure 13. Effect of Weather on Mean Driver Behavior Score

3.3.2.2 *Time of Day*

In previous studies, nighttime driving has been associated with more HRGCs accidents [1]. Figure 14 illustrates sample size and average driver behavior scores based on time of day. The figure reveals that drivers consistently received lower behavior scores during the nighttime traversals.

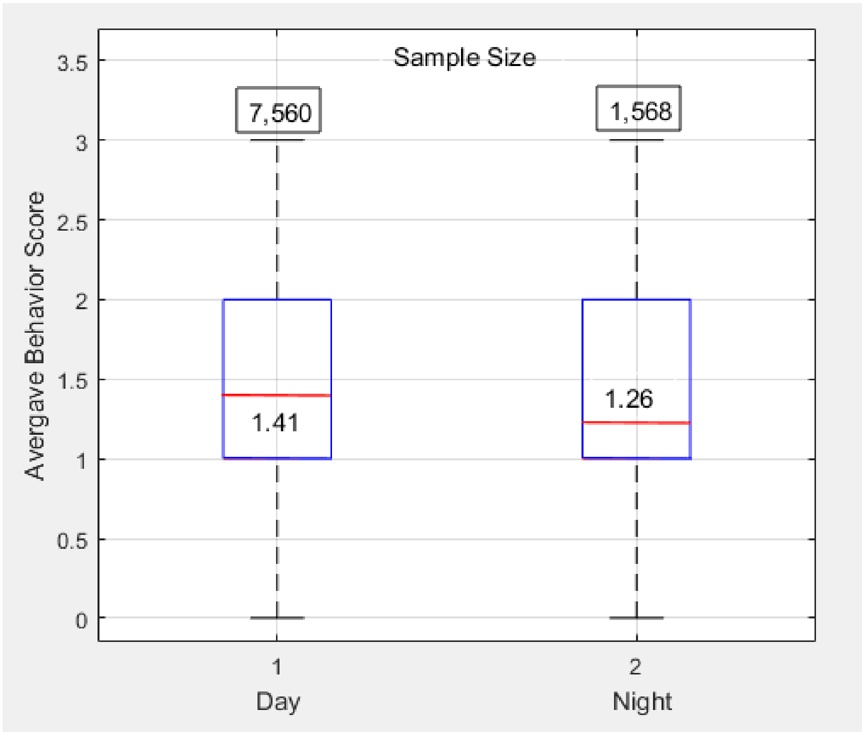


Figure 14. Drivers' Mean Behavior Score Based on Time of Day

3.3.2.3 Gender and Age

While driver demographics (gender and age) do not change between traversals, previous studies reported that male and younger drivers were involved in more HRGCs accidents compared to female and middle-aged drivers [1]. The results of this study, however, did not show any significant difference between the average driver behavior scores for male and female drivers (Figure 15).

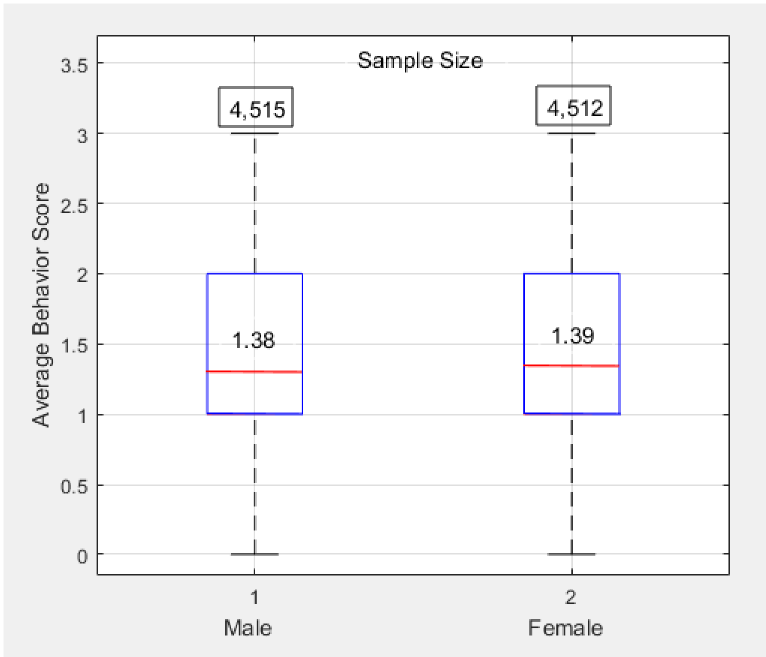


Figure 15. Mean Driver Behavior Scores by Gender

Driver demographic data were further grouped into three age categories to compare behavior scores between different age groups of male and female drivers: younger adults (16–34 years old), middle-aged adults (35–54 years old) and older adults (55 years old or higher). Figure 16 shows the sample size and drivers' average behavior scores based on gender and age groups. The data indicated that the difference in average behavior scores of male and female drivers in younger adults (1.4 versus 1.38) and older adults (1.39 versus 1.38) categories were negligible. The only noticeable difference was among middle-aged drivers, where female drivers received higher behavior scores compared to male drivers in the same age category (1.44 versus 1.34). Note that 139 records did not have a valid age record.

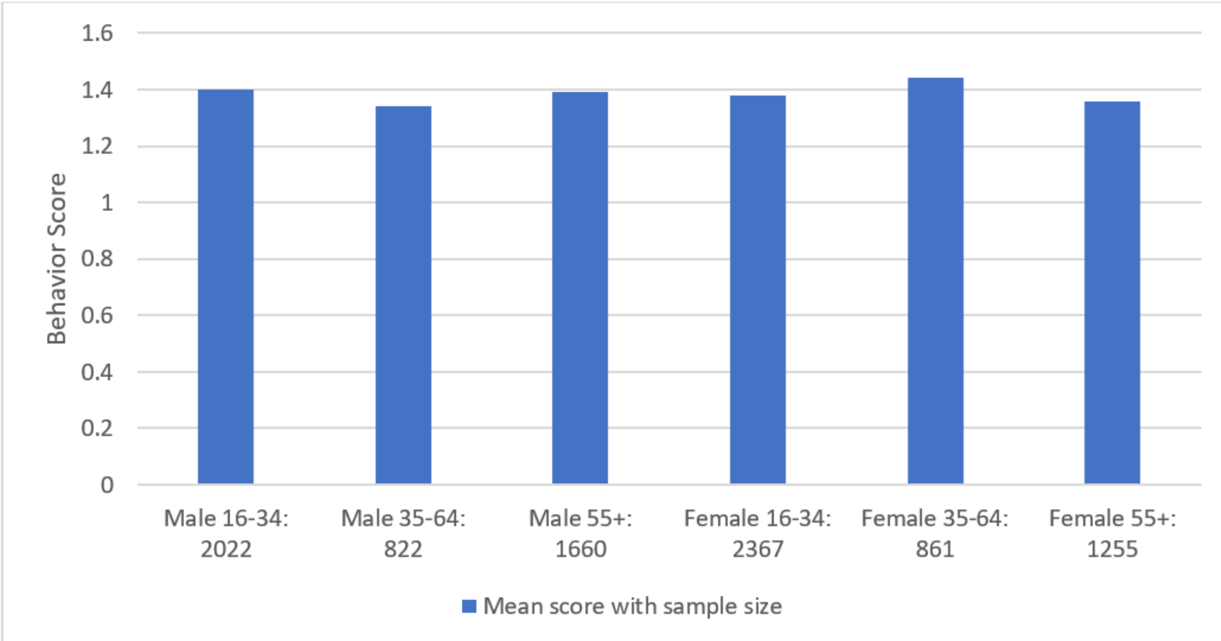


Figure 16. Mean Driver Behavior Scores by Age and Sex

3.3.3 Summary of Statistical Test Results for Weather, Time of Day, Gender, and Age

Table 11 shows the results of the t-tests conducted to compare parameter pairs under Category 2. Overall, several pairs (in bold letters) were found to have statistically significant differences in driver behavior score. Driver behavior under snowy conditions was significantly different from most other weather conditions (i.e., clear, cloudy, and rain). Perhaps the most consistent difference was between daytime and nighttime traversals, with daytime traversals seeing significantly higher mean scores.

Table 11. Summary of Statistical Test Results Based on Weather and Time of Day

Condition Pair	Deg. Freedom	t Stat	P(T<=t) two-tail	t Critical two-tail
WEATHER				
Clear vs. Cloudy	233	1.09	0.28	1.97
Clear vs. Rain	**472**	**2.70**	**0.01**	**1.97**
Clear vs. Snow	**92**	**-2.29**	**0.02**	**1.99**
Cloudy vs. Rain	432	0.66	0.51	1.97
Cloudy vs. Snow	**166**	**-2.54**	**0.01**	**1.97**
Rain vs. Snow	**127**	**-3.22**	**0.002**	**1.98**
TIME OF DAY				
Day vs. Night	**2,267**	**6.82**	**1.1E-11**	**1.65**

Researchers found no statistically significant difference between male and female drivers or between different age groups. However, when gender and age parameters were combined (Table 12), males in the 35–54 years old category had significantly lower average scores than females in same age group.

Table 12. Summary of Statistical Test Results Based on Gender and Age

Condition Pair	Age Group	Deg. Freedom	t Stat	p(T<=t) two-tail	t Critical two-tail
Male vs. Female	16-34	4330	0.79	0.43	1.96
	35-54	**1673**	**-2.32**	**0.02**	**1.96**
	55+	2662	0.41	0.68	1.96

4. Discussion and Limitations of Phase 1

The research team used a 3-point driver behavior score methodology developed at Michigan Tech and over 9,000 records of NDS data from the SHRP2 program to quantitatively evaluate driver behavior at HRGCs. Both the use of NDS data for the analysis and scoring methodology were novel and based on drivers' visual scanning and speed adjustment while approaching HRGCs. A higher behavior score was considered an indication of a more defensive driving behavior. Key Phase 1 findings are summarized in the following sections.

4.1 TCD Analysis

The results from the TCD analysis revealed no statistically significant difference in driver behavior at HRGCs with various types of TCDs, except for HRGCs with passive warnings that included stop signs. In general, the total driver behavior scores in HRGCs with passive warning devices aligned closely with those equipped with active warning devices, even if the visual scanning and speed adjustment components of the score were investigated separately. The average scanning scores of 1.13 suggests that drivers were scanning in both directions during approximately 60 percent of the traversals, while the speed score of approximately 0.284 suggests that drivers were preparing to stop at 30 percent of the HRGCs. A closer look at the actual scores revealed that in approximately 34 percent of the traversals, drivers looked both ways, 48 percent looked one way, and 18 percent failed to look at all. These values were consistent across all types of TCDs, except HRGCs with stop signs.

Scores for HRGCs with active TCDs showed a speed reduction approximately 24 percent of the time, while the corresponding percentage at passive HRGCs with crossbuck warnings only was lower (20 percent). The results were slightly better for HRGCs with yield signs – 32 percent of drivers prepared to stop; however, this percentage was based on only seven HRGCs.

The equal or even lower scores for visual scanning/speed adjustment at passive HRGCs is concerning. While locations with active TCDs provide drivers a visual warning of approaching trains, at passive HRGCs, visual detection is the only way for an individual to recognize potential danger. The majority of drivers who do not prepare for potential train arrival place themselves at risk during every traversal, providing one explanation for higher accident rates over time at passive HRGCs.

The only statistically significant exception were higher mean scores at passive HRGSs equipped with stop signs. The overall difference was mainly caused by better performance in the speed reduction category, as the percentage of drivers preparing to stop increased to above 70 percent when a stop sign was present. This percentage might have been even higher, as in some traversals a low entry speed, combined with failure to come to a complete stop, led to an insufficient speed reduction to score a point. The research team interpreted this result as confirmation that drivers expressed a higher level of defensive driving at HRGCs with stop signs. As such, the requirements in the MUTCD to increase their usage [19] seem warranted. On the other hand, it remains to be seen how permanent the improvement in driver behavior is – once drivers start to encounter them more frequently.

4.2 Accident Analysis

Fifteen HRGCs in the NDS database had accidents between 2000 and 2010. Researchers found that mean driver behavior scores at accident locations were similar to corresponding values in the full sample of HRGCs. The only notable difference was the higher mean score at the one (and only) passive HRGC with a yield sign; but this also highlights the shortcoming of the analysis, i.e., the small number of HRGCs in the accident dataset. Overall, due to the small sample size, the results were not generalizable to the general population; no conclusions could be made to differentiate driver behavior between HRGCs with and without accidents.

4.3 Trending Analysis

The initial trending analyses were based on looking at parameters individually and their impact on the behavior score. The analyses were broken down into two categories: Category 1, that concentrated on crossing-specific parameters, and Category 2, that concentrated on traversal-specific parameters. For Category 1, the analyses were exploratory in nature and not tested for statistical significance.

4.3.1 Category 1 – Crossing-Specific Parameters

The trending analysis on crossing-specific parameters showed several linear trends across the various values. While the statistical significance was not tested, the results can be considered at least as valuable observations that include the following:

- The preliminary trend analysis (Figure 12) revealed a moderate decrease in both the scanning and speed scores as AADT increased. One explanation may be that as AADT increases, the use of active TCDs can also be expected to increase. The added reliance on the active TCDs is likely to reduce the mean driver behavior scores. In addition, with higher AADT drivers are more likely to follow the traffic flow when traversing a HRGC.
- Scanning behavior appeared to increase slightly as the number of trains per day (TPD) increased (Figure 13). Interestingly, the scan scores for HRGCs with no reported through trains were very similar to the rest of the HRGCs in the database. Speed scores did appear to increase with train traffic, suggesting that drivers were more likely to reduce their speed with more frequent train traffic.
- There was a trend for both lower scanning and speed behavior scores as the highway speed limit increased (Figure 14). Just like with AADT, the research team believed this decline related to the increased presence of active TCDs on roadways with higher speed limits. However, the results could also suggest that at higher speeds, driver attention was more focused on the features within the roadway, leaving less attention for potential trains.
- Scanning behavior did not seem affected by the train speed at HRGCs (Figure 15). However, the speed scores increased as train speed increased, suggesting that drivers were more aware of the potential danger from higher-speed trains.

4.3.2 Category 2 – Traversal-Specific Parameters

While the Category 2 analysis did not reveal major trends between the scenarios, the research team found statistically significant differences among several conditions:

- The team found that under different **weather** conditions, drivers received the highest behavior scores in snow and the lowest in rainy conditions (fog was even lower, but the sample size was too small for statistical analysis). Since a low driver behavior score was indicative of less defensive driving, the results support the findings from previous studies for increased risk in rainy conditions. However, they did not support the previous finding for increased accident risk under snowy conditions.
- The results were more consistent on differences in driver behavior based on **time of day.** The team found that all drivers from all gender/age groups received significantly lower behavior scores during the night compared with day. This outcome supports previous studies that have revealed poor visibility conditions to have negative impact on driver behavior, thus increasing accident risk at HRGCs [15]. Based on this study's results, the team suspected that poor nighttime visibility led to higher driver concentration on the road ahead, and as such, to lower driver behavior scores.
- Researchers did not find any significant difference in average behavior scores of **male and female drivers.** Based on driver demographic analysis, the only statistically significant difference was between **middle-aged** female drivers (35-54 years old) when compared to male drivers in the same age category. This differs from earlier studies that concentrated on accidents at HRGCs and found younger male drivers to be at higher risk.

5. Phase 2: Data Sources and Methodology for Driving Simulator Study

The research team selected two HRGC locations (crossing ID 521090P and 621549W) from the RID database for replication in the driving simulator. These crossings were selected based on the number of traversals available in the NDS data and the expected level of difficulty for modeling the specific HRGC environment in the simulator. The HRGCs were virtually modeled in collaboration with the National Advanced Driving Simulator (NADS) team at Iowa State University, using NADS MiniSim^{TM} software (Appendix B).

The first HRGC (scenario A) featured active TCDs (gates and flashing lights), and the second HRGC (scenario C) featured cantilevered flashing lights only as the active TCD. The simulator studies were conducted in two batches, but the main analysis combined studies one and two into a single “SIM” dataset.[1]

Table 13 presents a summary of parameters for both the NDS and simulator analysis. Driver behavior was quantified using the same driver behavior score as in the NDS data analysis. Since the team had access to the un-blurred video of each participants face in the SIM dataset, visual scanning behavior was coded manually, as opposed to automatic classification based on estimated head tracking data.

Table 13 - Description of NDS vs. Simulator Data

	NDS	**SIM**
Speed limit	45 mph	Study 1 – no limit
		Study 2 – 45 mph
# of Drivers	24	48 (in analysis)
# of Traversals	284 (total) Scenario A (gate/lights) – 51 Scenario C (lights only) – 233	256 (total) Scenario A (gate/lights) – 124 Scenario C (lights only) – 132
Advanced Warning Distance	Determined from videos and Google Earth	(same as NDS)
Source of head rotation data	Automated head tracking	Manual video analysis

[1] Researchers first analyzed simulator study 1 and study 2 individually. There was no significant difference in each study, so researchers combined the results into a single "SIM" dataset for main analysis. If the comparison between NDS data and each SIM dataset showed distinct difference, researchers separated the results and provided each analysis results separately (Figures 24, 25 & Table 15). However, even those differences came from relative comparisons between NDS and each study and proved not to be statistically significant. Individual Study 1 results can be found at Landry, S., Wang, Y., Lautala, P., Nelson, D., & Jeon, M. (2018, July). Driver Behavior at Simulated Railroad Crossings. In *International Conference on Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management* (pp. 599-609). Springer, Cham.

	NDS	**SIM**
Source of speed analysis data	NDS vehicle data	SIM vehicle data

A total of 51 traversals were included in the NDS dataset for scenario A (gates/lights), and 233 traversals were included in the NDS dataset for scenario C (lights only). This dataset included 24 unique drivers (M_{age} = 41.8, SD_{age} = 23.5 13 male, 11 female). For the simulator study, a total of 54 participants (M_{age} = 20.12; SD_{age} = 1.71; 40 male, 14 female) were recruited from Michigan Tech's undergraduate psychology courses. The participant sample had an average of 4.83 years (SD_{years} = 1.71) of driving experience.

Researchers collected data in two closely aligned studies. Study 1 included 20 participants, and study 2 included the remaining 34. Participants were not informed about the specific goals of the study beforehand, as the recruitment advertisements only mentioned that the experiment was about "driver behavior in a medium fidelity simulator."

The final dataset for study 1 included 18 participants, and for study 2 included 30. The remaining six participants were dropped from later analysis due to missing data caused by experimenter error or technical difficulties. The final SIM dataset for the analysis included 124 simulated traversals at the crossing with lights/gates (scenario A), and 132 traversals at the crossing protected by lights only (scenario C).

All experimental stimuli and protocol were identical in each simulator study. The instructions for participants were the only difference. In study 1, participants were instructed to ignore the simulated speedometer and only use speed perception cues from the simulated environment when determining the appropriate speed to drive; this was due to the pilot participant feedback that suggested a mismatch between the visual cues of motion and the simulator speedometer. In study 2, participants were instructed to obey a 45-mph speed limit throughout the session.

5.1.1 Experimental Design

The research team used two existing HRGCs in a rural setting (included in the NDS dataset) to develop corresponding simulated scenarios (Figure 17). Each scenario was included in a loop that participants drove around three times in succession. Each lap included two HRGCs, resulting in six traversals per scenario for each participant. Study 1 included a train present event on the final (sixth) traversal for both scenarios, but otherwise both studies had identical settings (excluding the speedometer instructions). The crossings were equipped with lights and gates (Figure 18), and a HRGC with flashing lights in cantilevered support (Figure 19). The order of scenario exposure was counterbalanced across participants, meaning half the participants experienced scenario A first, and the other half experienced scenario C first. Each scenario was around 20 minutes in length, depending on the speed of the participant.

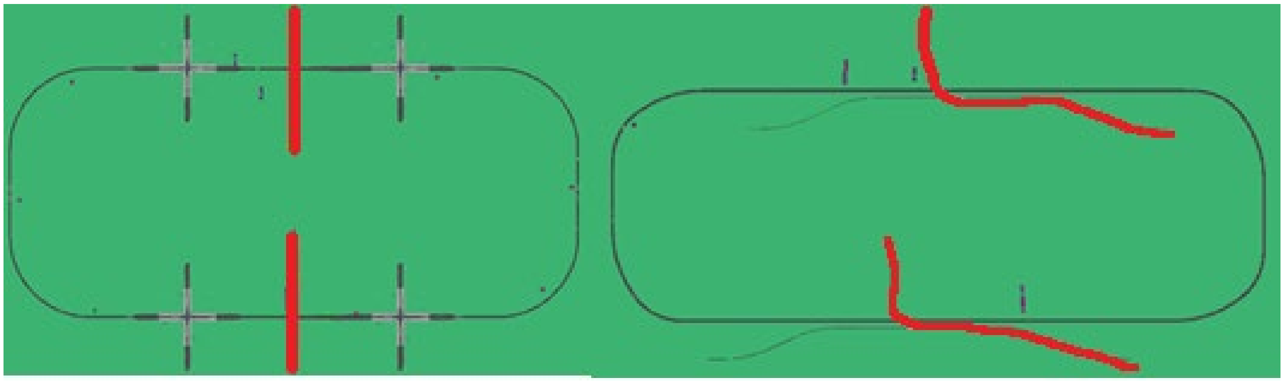


Figure 17. Track Designs for Scenario A (Left) and C (Right). Red Lines Indicate Railroads that Intersect with Highways (Gray)

Figure 18. Simulated Advance Warning (Left) and Gated Crossing (Right) in Scenario A

Figure 19. Simulated Advance Warning in Scenario C (Left) Compared to Reference NDS Video (Right)

After completing both scenarios, participants answered the following questions:

1. How many years have you been driving?
2. How realistic were the scenarios?
3. How different is your real-world driving behavior from your simulated driving behavior?

4. How many times do you encounter a train crossing per month (on average)?
5. How many times have you had to come to a stop for a train at a railroad crossing in your lifetime?
6. What would/should you do when encountering this sign? (The question would present a picture of each of the different HRGC TCDs featured in the experiment.)
7. How did your behavior at railroad crossings change over the course of this experiment?
8. How noticeable was the head tracking device and cap? Do you think it had any impact on your behavior?

5.2 Results and Data Analysis

First, we explored the impact of the different instructions between SIM Study 1 and SIM Study 2, and their differences compared to the NDS dataset (Figure 20). A post-hoc Tukey honest significant difference (HSD) test for multiple comparisons was run to investigate the differences between each subgroup (Table 14). The post-hoc Tukey HSD runs multiple unpaired t-tests simultaneously while controlling for the additional family-wise type 1 error rate (false positives due to multiple comparisons). The results suggested that NDS behavior scores were significantly higher than SIM scores in study 1 while the difference between NDS and study 2 scores fell just short of the significance criteria ($p = 0.056$,). There was no significant difference between the two simulator studies' mean behavior scores ($p = 0.831$). Based on this null result, SIM studies 1 and 2 were combined for the remainder of the analysis as a single "SIM" dataset. Figure 21 depicts the means and standard errors for each subgroup to visualize the interaction between data source and TCD type.

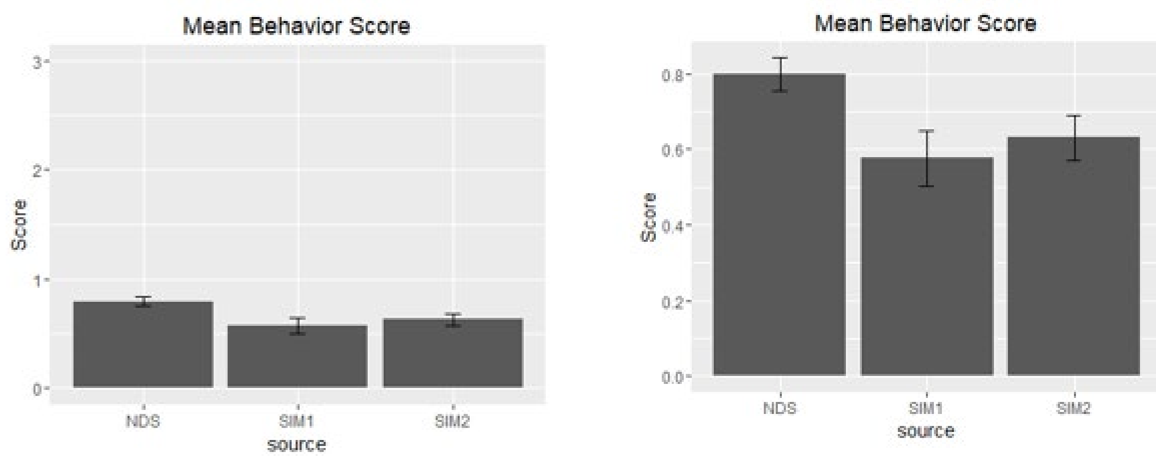


Figure 20. Mean Behavior Score Grouped by NDS and SIM (Study 1 and Study 2) Datasets

Table 14. Post-Hoc Tukey HSD Analysis

Subgroups	Diff	lwr	upr	p adj
SIM1-NDS	-0.223	-0.425	-0.021	0.026*
SIM2-NDS	-0.168	-0.341	0.003	0.056
SIM2-SIM1	0.054	-0.167	0.277	0.831

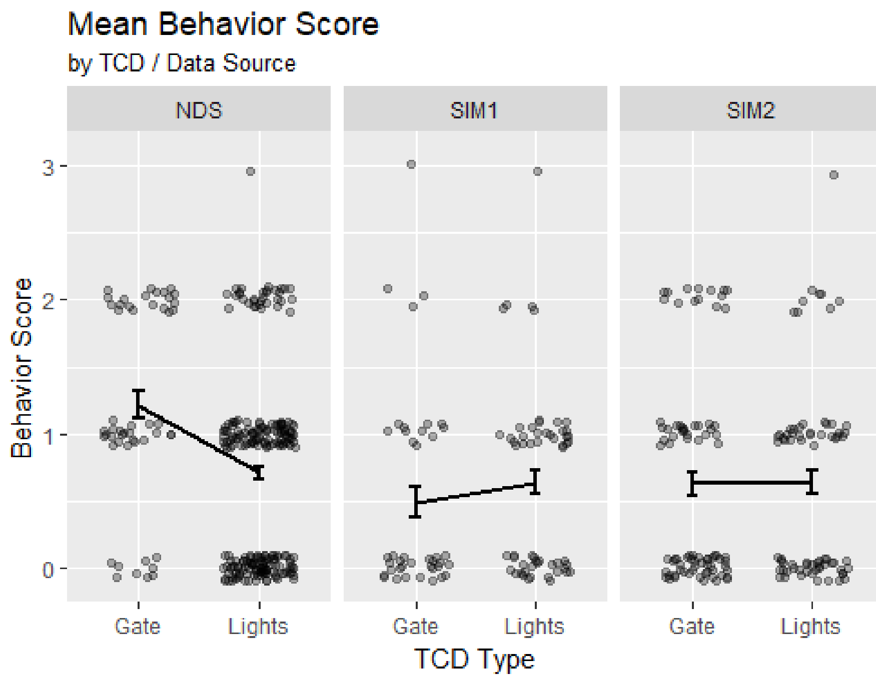


Figure 21. Mean Behavior Score Grouped by NDS, Study 1, and Study 2 Datasets

Next, the research team conducted a 2x3 analysis of variance (Table 15) to detect significant differences between simulated and natural data sources (SIM vs. NDS) and TCD types (gate/lights vs. lights only) on driver behavior scores. Groups were organized by data source (simulated vs. naturalistic datasets) and TCD type (gate/lights (A) vs. lights only (C) scenario). Table 15 reveals that the p-value for each was below 0.05, suggesting that significant main effects on driver behavior scores existed for both TCD types and data sources, as well as for their interaction.

Table 155. 2x3 Analysis of Variance

Category	**DF**	**Sum Sq**	**Mean Sq**	**F value**	**p<0.5**
Data Source	2	5.04	2.519	4.788	0.008*
Scenario (TCD Type)	1	2.85	2.848	5.413	0.02*
Source: Scenario	2	8.47	4.237	8.052	0.0003*
Residuals	534	280.99	0.526		

The team conducted an independent sample t-test to compare total driver behavior scores between the SIM and NDS datasets. Results suggested driver behavior scores were significantly higher in the NDS dataset than the SIM dataset ($t(533) = 2.99$, $p = 0.003$). Figure 22 depicts mean behavior score across datasets. The error bars show the standard error of the mean. The left graph y-axis is scaled to show the full 3-point range of the behavior score. The right graph depicts the same data, only zoomed in for easier visual comparison. A similar method is used for the following graphs as well.

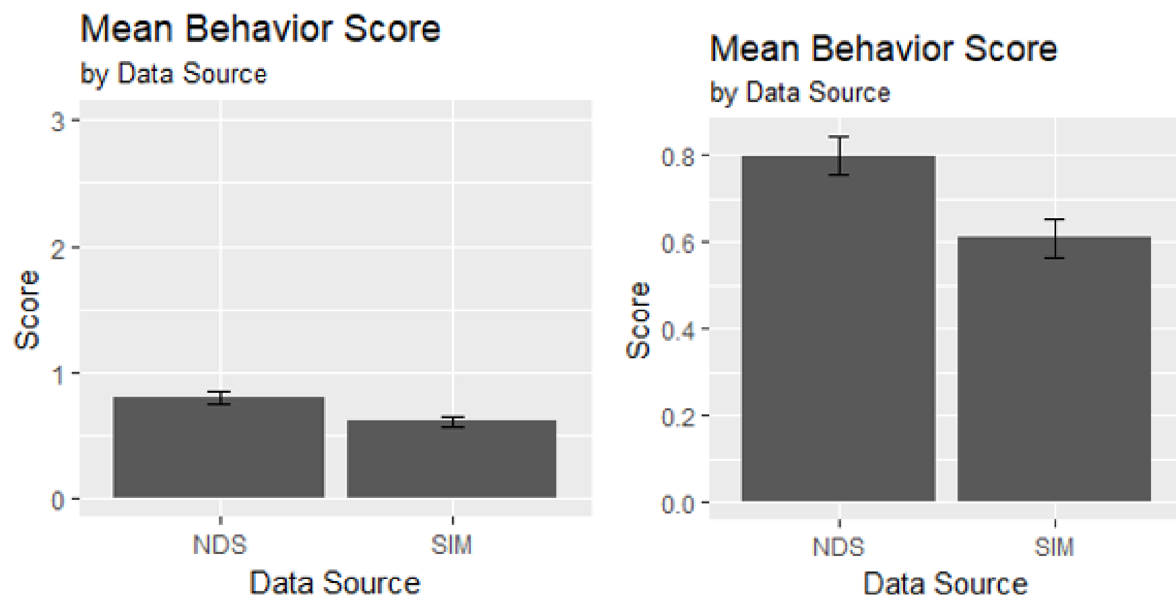


Figure 22. Mean Behavior Score by Data Source (NDS vs. SIM)

Researchers conducted a second independent sample t-test to compare the behavior scores between the two types of TCDs (scenario A vs. C). Results suggested driver behavior scores were not significantly different between the scenarios across both NDS and SIM datasets ($t(310) = 1.17$, $p = 0.24$). Figure 23 depicts mean behavior scores for both scenarios (types of TCD).

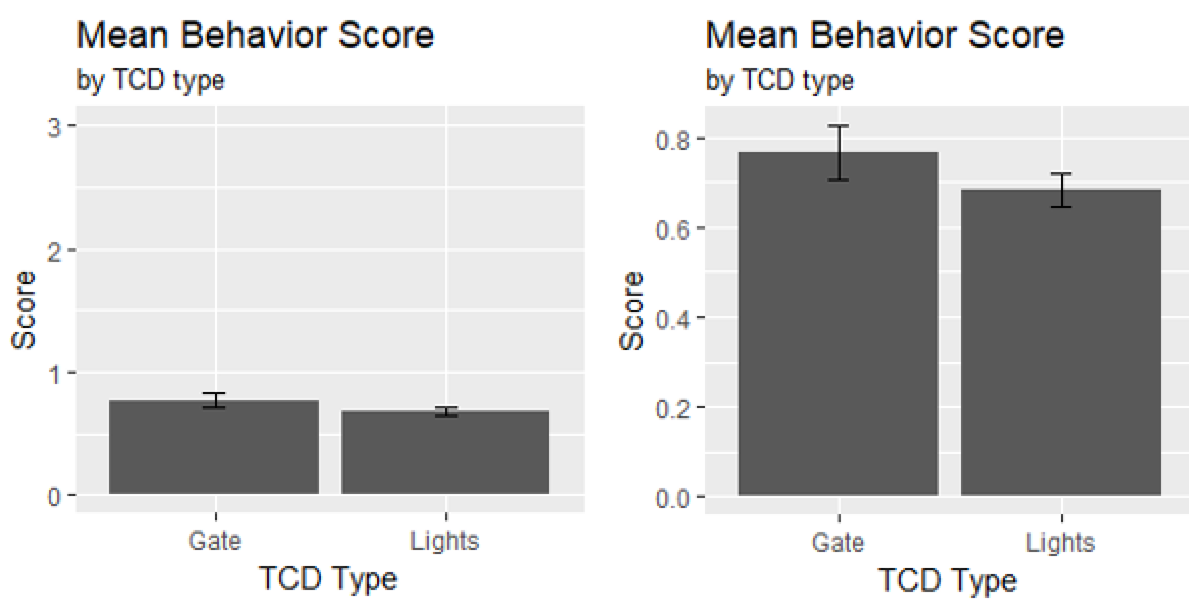


Figure 23. Mean Behavior Score by Scenario (TCD Type, Scenario A vs. C)

A significant interaction effect suggests the effect of TCD type was larger in the NDS dataset than for the SIM dataset (Figure 24). When only considering SIM data, there was no statistical difference between scenario A and C. However, when only considering NDS data, driver behavior scores were higher in response to scenario A (gates/lights) than for scenario C (lights only). For the interaction plot in Figure 25, the small dots represent individual data points with artificial jitter (right). A significant interaction difference is indicated by the slopes between the two data sources when comparing responses to different TCD types.

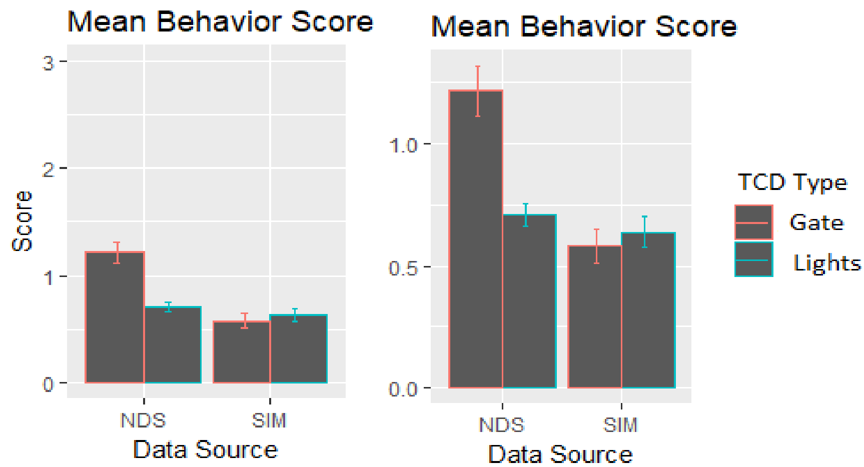


Figure 24. Mean Behavior Score by Data Source for Scenarios A and C

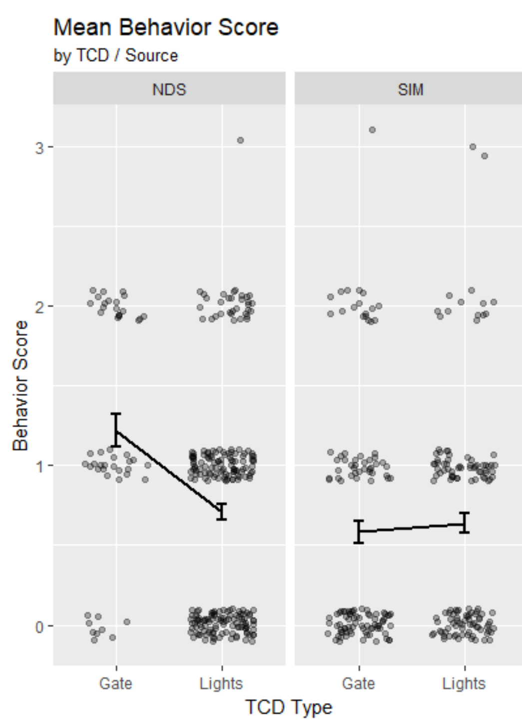


Figure 25. Interaction Plot by Data Source for Scenarios A and C

5.2.1 Summary of Behavior Scores

Figure 26 is a stacked bar chart depicting the contribution of each behavior score component (visual scanning vs. speed reduction) to the total driver behavior score. The yellow portion represents the mean speed reduction score (total 1 point possible), and the gray portion represents the mean visual glance score (total 2 points possible). In both datasets (NDS and SIM), most participants' scores were from visual glances, and very few were from speed reduction. Only 6.3 percent (18/284) of traversals received a speed reduction point in the NDS dataset, and only 3.1 percent (8/256) of traversals received a speed reduction point in the SIM dataset (Table 13).

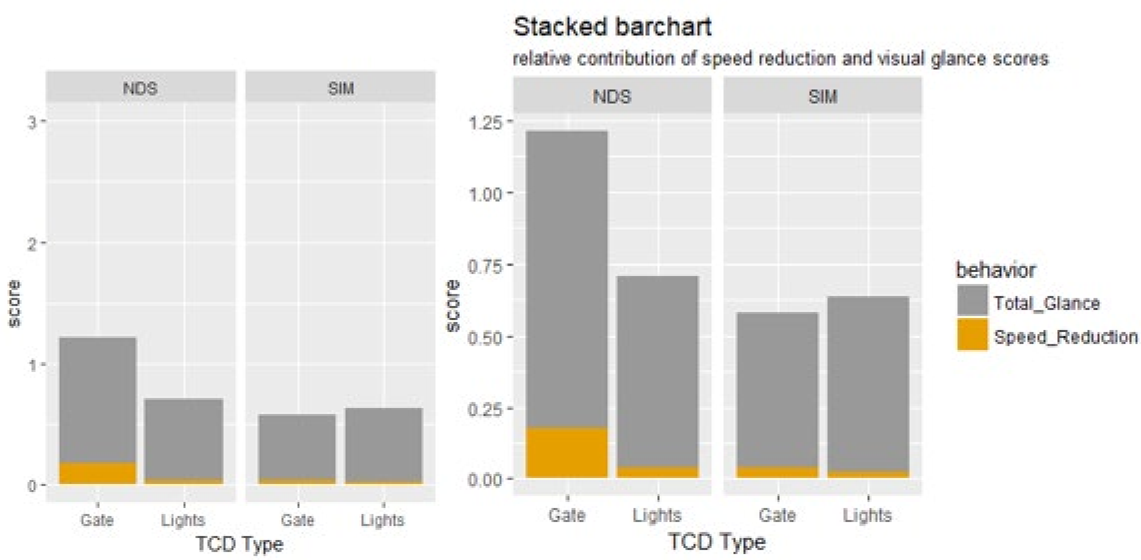


Figure 26. Stacked Bar Chart Depicting the Contribution from Behavior Score Sub-Behaviors

Table 16 provides the percentage distribution of the four possible driver behavior scores (0, 1, 2, 3). The score patterns were different for NDS and SIM datasets for scenario A (gates/lights). In the NDS dataset, a majority of drivers scored 1 or 2 points, while less than half did in the SIM dataset.

Researchers also observed a low number of drivers scoring a point from speed reduction in both datasets, 6.3 percent and 3.1 percent, respectively (Table 17). Hence, the majority of points came from visual scanning behavior.

Table 16. Percent of Total Behavior Score

TCD Type	Score Bin	NDS Count	NDS Percentage	SIM Count	SIM Percentage
Gate/Lights	0	9	17.65	71	57.26
-	1	22	43.14	98	28.23
-	2	20	39.22	32	13.71
-	3	0	0	1	0.81
Lights Only	0	102	43.78	64	48.48
-	1	98	42.06	54	40.91
-	2	32	13.73	12	9.09
-	3	1	0.43	2	1.52

Table 17. Count and Percent of Traversals with Speed Reduction during the Taversal

Speed Reduction Point Awarded	NDS	SIM
Fail	266	248
Pass	18	8
Pass Percent	6.30%	3.10%

5.2.2 Summary of Post-Experiment Survey

Key findings from the post-experiment survey include:

- All but one participant reported that the driving simulator scenarios were realistic. A few participants suggested the low visual quality and the unnatural response of the vehicle as two potential features that could detract from the realism of the simulation.
- Roughly half of the participants reported there was little to no difference between their driving style in the simulator and their driving style in the real world. Fourteen participants admitted that they drove less cautiously in the simulation due to the lack of consequences. Alternatively, seven participants admitted that they drove more cautiously in the simulation because it was an experiment where their behavior was being closely monitored.
- Participants where shown pictures of each of the TCDs and were asked to describe the appropriate driver response to each. All participants understood it was their responsibility to observe and yield to oncoming trains at HRGCs with passive TCDs. Approximately half of the participants reported they would approach active TCDs in the off position with caution by slowing down and scanning for trains. The other half suggested that active TCDs in the off position indicate drivers do not need to slow down or look for trains.

- Almost half of the participants experienced one to three HRGCs on a typical month of driving (Figure 27 – top), and almost half of the participants have come to a complete stop because of a train fewer than five times in their lifetime (Figure 27 – bottom).

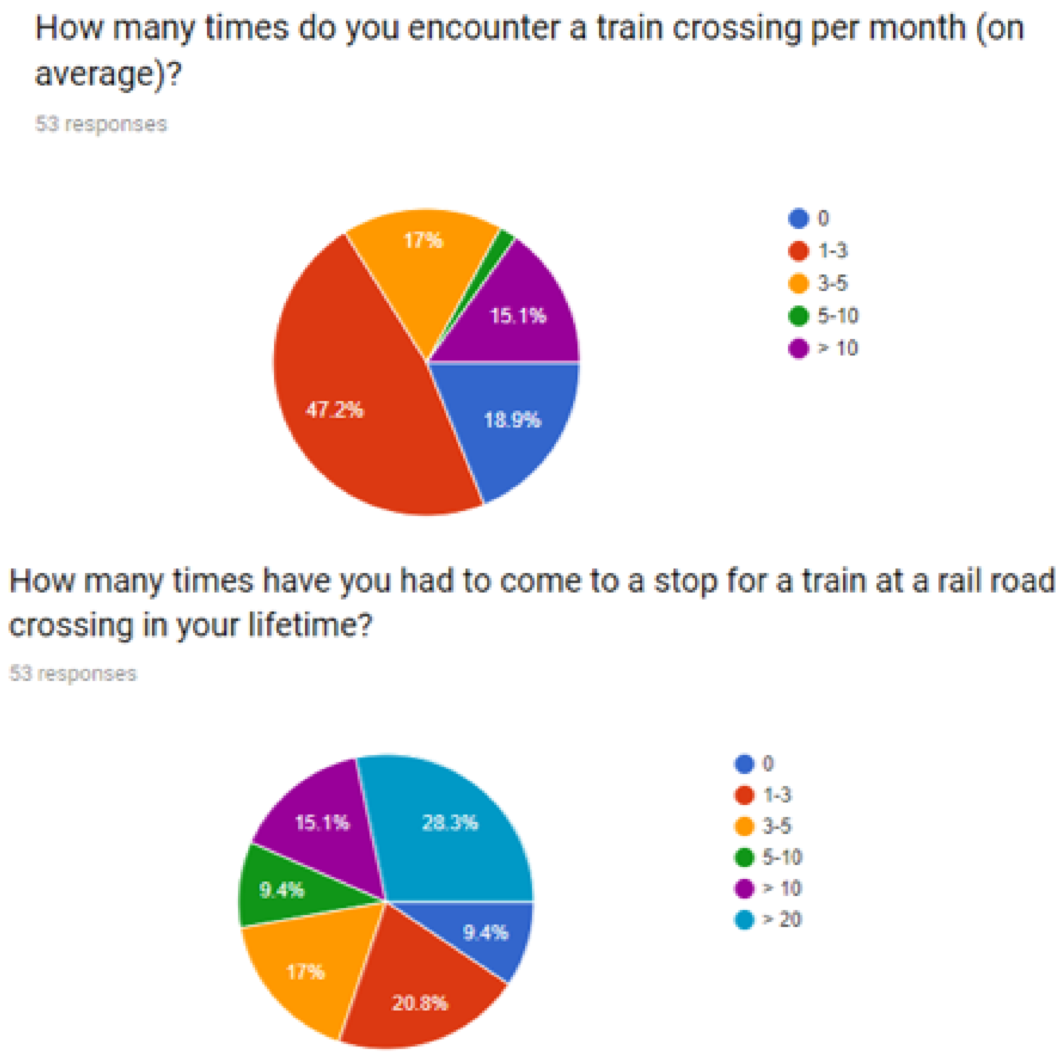


Figure 27. Pie Graphs of Exposure to Crossings and Trains in Real-World Driving

5.3 Discussion of Phase 2

Researchers used two existing HRGCs included in the NDS dataset to develop simulated scenarios in the NADS. They then recruited student drivers to investigate driver behavior at HRGCs in the simulated settings and to compare driver behavior between the NDS and simulated datasets. The evaluations used the same 3-point quantitative method as the NDS analysis.

Key findings from Phase 2 of the study include:

- A post-experiment survey revealed that the simulator was considered fairly realistic and participants maintained their driving style. However, the higher operating speeds by study 1 participants (and the post-experiment survey results) suggest that the speed perception in the simulator differed from that in the natural environment.
- Comparisons between SIM data from study 1 and study 2 highlight the effects of train presence and verbal speed instructions, as those were the only differences between the two studies. Participants drove around 80 mph in the simulator without explicit verbal instructions to follow the 45 mph speed limit. However, the researchers found no statistical difference for mean behaviors scores between stimulation studies 1 and 2, which suggests participants approached HRGCs with the same amount of caution regardless of average vehicle speed and the inclusion/exclusion of train present events.

The main findings from comparing the results of NDS and SIM datasets are as follows:

- Both datasets suggest that drivers' behavior at the crossings have room for improvement from a visual scanning and speed adjustment perspective. Mean driver behavior scores were low for both NDS and SIM datasets, 0.8 and 0.6 out of possible 3 total points, respectively. Speed reduction/adjustment behavior was poor in both NDS and SIM datasets. Very few crossing events (26/540, or 4.8 percent of total) in either dataset (NDS and SIM) were awarded a point for speed reduction using current criteria and very few drivers (4/540, or 0.7 percent of total) received the full 3 points for the driver behavior score.
- A closer look reveals the difference between the two datasets. Statistical tests (ANOVA and t-test) suggested a statistically significant difference in the driver behavior scores between NDS and SIM datasets, but not between the different TCDs (scenario A vs. C).
- The significant main effect for data source suggests participants in the simulation approached HRGCs less cautiously than participants in the NDS dataset. The significant main effect for TCD type suggests drivers approached HRGC with lights/gates more cautiously than HRGC with flashing lights only – but researchers only observed this trend in the NDS dataset in post-hoc subgroup analyses. Similarly, behavior scores were higher for scenario A (gates/lights) in the NDS dataset, but the same trend was not observed in the SIM dataset.
- Based on the post-experiment questionnaire, some participants drove in the simulator more cautiously than in their actual driving, while other participants drove less cautiously. Also, half of them reported they would approach active warnings in the off position with caution, but the other half reported they would not slow down or scan for trains. All participants understood it was their responsibility to search for and yield to oncoming trains for passive railroad warnings, but (consistent with other driving behavior research) their subjective responses did not correspond with their actual behavior patterns.

6. Study Limitations

This study was the first attempt to use the SHRP2 NDS data for the driver behavior analysis at the HRGCs. The research team developed a novel "driver behavior score" methodology for the analysis and for the comparisons with simulator experiments conducted in Phase 2. The team wants to highlight the following limitations and challenges identified in its data and approach during the study.

- **Completeness and accuracy of NDS data:** Due to the difficulties with keeping the sensor arrays in 3,500 vehicles running all the time over an extended period of time, some of the NDS data records had missing or incomplete data. For example, some data records missed GPS data and numerous records didn't have data for gas or brake pedal depressions. The team investigated methods for circumventing the missing data and developed some techniques to bring some of the excluded records back into the analysis.
- **Use of a three-point score as the sole qualifier for the analysis:** For simplicity, the methodology relied on developing a single driver behavior score, based on two activities (head rotation and speed reduction) as a quantitative indicator for all behavior during the HRGC traversal. However, condensing a whole chain of events into a single score limits the possibility of investigating the impact of specific factors on driver behavior. It also combines large number of HRGCs with varying characteristics and excludes certain types of HRGCs from the analysis (such as HRGCs near highway intersections). An alternative way to use NDS data in HRGC safety analysis is concentrating on a single behavior, such as the location of speed reduction. Such targeted analysis might provide more granular data for parametric safety analysis.
- **Limited sample size:** The analysis contained over 9,000 samples, a significant amount of data for the analysis. Dividing the 9,000 samples into subcategories resulted in some subcategories having especially small sample sizes (e.g., HRGCs with passive TCDs), which limited the researchers' ability to conduct comparative analyses. This is also true for the simulator research. Two HRGCs and 40+ candidates with limited background diversity are at a best a good start for statistical analysis.
- **Simulator perception:** Research has shown that the level of realism or the fidelity of the driving simulator scenario will not necessarily influence the study outcomes. However, the research team acknowledges that the perception about the actual risk may be different between the two and can affect the study outcomes.

7. Conclusion and Recommendations for Future Work

In Phase 1 of this study, the researchers used over 9,000 individual traversals obtained from the SHRP 2 NDS data to evaluate driver behavior at HRGCs. They developed a 3-point behavior score based on visual scanning for trains and vehicle speed reduction for the quantitative evaluation. Mean scores were used to generate comparisons and trends, based on selected parameters and conditions. The main comparisons included HRGCs divided by different TCDs and with or without an accident history. In Phase 2, they implemented selected HRGCs from an NDS dataset in a driving simulator. They used the same evaluation methodology to compare and contrast the simulator results with NDS ones.

While the NDS data analysis resulted in numerous interesting observations, they showed little statistical difference in driving behavior among any of the TCDs analyzed. The only exception was the significantly higher mean scores at passive HRGCs equipped with stop signs. The other consistent finding was the higher mean scores for traversals during the day versus nighttime. Researchers found no significant evidence of systematic driver behavior differences in most other categories tested, such as behavior between genders or between age groups.

It was evident from the NDS data (and driver behavior score) that most drivers were not scanning for trains, nor were they preparing to stop, even at crossings with passive warning devices where drivers must rely on their own observations for a safe passage. For example, in the driving simulator scenarios, less than 10 percent of all participants (NDS and simulator data combined) received a point for a proper speed adjustment. The researchers plans to conduct further research on how to encourage drivers to behave appropriately using additional effective but cost-efficient methods (e.g., in-vehicle alerts).

In Phase 2, researchers used a driving simulator experiment and related comparisons to investigate similarities and differences in the driver behavior scores between simulated and natural settings. In general, similar trends could be observed in each dataset, but there were more significant differences in scores between scenarios in the NDS than in the simulated setting, a finding perhaps attributed to participants' perceptions that the simulated environment is safe.

A better understanding of driver behavior at HRGCs would help predict situations where drivers are less cautious and could be at risk of accidents. Despite the limitations and shortcomings of the current effort, the NDS data can help improve our understanding of driver behavior at HRGs. The team also believes that driving simulators allow researchers to quickly and efficiently analyze new methods to address driver behavior and evaluate potential safety improvements at HRGCs.

The methodologies and data processes developed in this study to evaluate parameters, such as weather, demographics, time of day, etc., offer an opportunity to quantify the performance of everyday drivers on a large scale. The limitations identified in this report should be addressed and its methodologies modified to improve the accuracy and credibility of the analysis. Instead of relying on a single 3-point scale to evaluate driver behavior, the next steps in the research may concentrate on improving the understanding of individual parameters, such as the exact locations where drivers adjust their speed when approaching a HRGC. Researchers may also consider the use of multivariate analysis techniques to investigate which environmental variables, or groups of variables, have significant effects on driver behavior at HRGCs. Future research can also explore applications of machine learning and artificial intelligence for predicting driver behavior and the perceptions of risks.

8. References

1. Federal Railroad Administration. (2017). In-Depth Data Analysis of Grade Crossing Accidents Resulting in Injuries and Fatalities [DOT/FRA/ORD-17/04]. Washington, DC: U.S. Department of Transportation.
2. Sheikh, Y., et al. (2004). Visual Monitoring of Railroad Grade Crossing. *Proc. of SPIE*.
3. Federal Railroad Administration. Grade Crossing Data Tool.
4. Federal Railroad Administration. (2016). Analysis of Grade Crossing Accidents Resulting in Injuries and Fatalities [RR 16-10]. Washington, DC: U.S. Department of Transportation.
5. Lerner N., Ratte D., & Walker J. (1990). Driver Behavior at Railway-Highway Crossings [FHWA-SA-90-008]. Washington, DC: U.S. Department of Transportation, Federal Highway Administration.
6. Federal Railroad Administration. (2008). Driver Behavior at Highway-Railroad Grade Crossings: A Literature Review from 1990-2006 [DOT/FRA/ORD-08/03]. Washington, DC: U.S. Department of Transportation.
7. Federal Railroad Administration. (2013). Driver Behavior Analysis at Highway-Rail Grade Crossings using Field Operational Test Data—Light Vehicles [DOT/FRA/ORD-13/28]. Washington, DC: U.S. Department of Transportation.
8. Strategic Highway Research Program 2 (SHRP2). Naturalistic Driving Study – Insight Data Access Website.
9. Iowa State University, Center for Transportation Research and Education. SHRP 2 – Roadway Information Datatbase.
10. Federal Railroad Administration. Highway-Rail Crossing Inventory Data.
11. Federal Railroad Administration. Crossing Accident /Incident Data.
12. Federal Railroad Administration. Highway-Rail Grade Crossings Overview.
13. Muhire, M., Lautala, P., Nelson, D., & Dean, A. (2017). Selection of Representative Crossings Database for the Evaluation of Driver Behavior Over Highway-Rail Grade Crossings. ASME/IEEE Joint Rail Conference, 2017 Joint Rail Conference [doi:10.1115/JRC2017-2294].
14. Dean, A., Lautala, P., & Nelson, D. (2017). Effectiveness of Using SHRP2 Naturalistic Driving Study Data to Analyze Driver Behavior at Highway-Rail Grade Crossings. ASME/IEEE Joint Rail Conference, 2017 Joint Rail Conference [doi:10.1115/JRC2017-2288].
15. Dean, A., Nelson, D., Lautala, P., & Jeon, M. (2017). Development and Validation of Post-Processing Mehods for the SHRP2 MASK Head Pose Data. In *Transportation Research Circular E-C229: 10th SHRP2 Safety Data Symposium*, Washington, DC, October 6, 2017 [ISSN 0097-8515].
16. Lautala, P., Nelson, D., Jeon M., & Muhire, M. (2017). Using NDS Data to Evaluate Driver Behavior at Highway–Rail Grade Crossings. In *Transportation Research Circular E-C229: 10th SHRP2 Safety Data Symposium*, Washington, DC, October 6, 2017 [ISSN 0097-8515].

17. Lautala, P., Muhire, M., Salim, A., Jeon, M., Nelson, D. & Dean, A. (2018). Assessment of Driver Compliance at Highway–Rail Grade Crossings Based on Naturalistic Driving Study Data [TRB 18-06234]. *Proceedings of the TRB 97th Annual Meeting*. Washington, DC, January 7-11, 2018.

18. Salim, A. (2018). Evaluation of Driver Behavior at Highway-Railroad Grade Crossings Based On Environmental Conditions And Driver Demographics. MS Report, Michigan Technological University.

19. Federal Railroad Administration. (2009). Manual on Uniform Traffic Control Devices (MUTCD), Part 8, Traffic Control for Railroad and Light Rail Transit Grade Crossings, 2009.

Appendix A. - Percent Behavior Scores

Table 18 presents an analysis showing the percentage of drivers who received each level of behavior score. The table shows total score as well as scan and speed scores separately. Figure 28 through Figure 30 show the same comparisons in a graphic format.

Table 18. Behavior Score by Percent

Scan Score	# Traversals	Percent	Speed Score	# Scores	Percent	Total Score	# Scores	Percent
All Traversals								
2	3071	33.6	1	2135	23.4	3	780	8.5
1	4382	48.0	0	6993	76.6	2	3320	36.4
0	1675	18.4		9128	100.0	1	3679	40.3
Total	*9128*					0	1349	14.8
Crossbuck with Yield								
2	75	34.2	1	69	31.5	3	33	15.1
1	107	48.9	0	150	68.5	2	72	32.9
0	37	16.9		219	1.000	1	83	37.9
Total	*219*					0	31	14.2
Crossbuck Only								
2	187	35.0	1	105	19.6	3	42	7.9
1	257	48.0	0	430	80.4	2	197	36.8
0	91	17.0		535	1.000	1	216	40.4
Total	535					0	80	15.0
Lights, No Gates								
2	456	32.9	1	331	23.9	3	129	9.3
1	674	48.6	0	1056	76.1	2	474	34.2
0	257	18.5				1	582	42.0
Total	*1387*					0	202	14.6
Gates								
2	2353	33.7	1	1630	23.3	3	576	8.2
1	3344	47.9	0	5357	76.7	2	2577	36.9
0	1290	18.5				1	2798	40.0
Total	*6987*					0	1036	14.8
Crossbuck with Stop								
2	73	44.8	1	125	72.3	3	52	31.9
1	75	46.0	0	48	27.7	2	73	44.8
0	15	9.2				1	37	22.7
Total	*163*					0	1	0.6

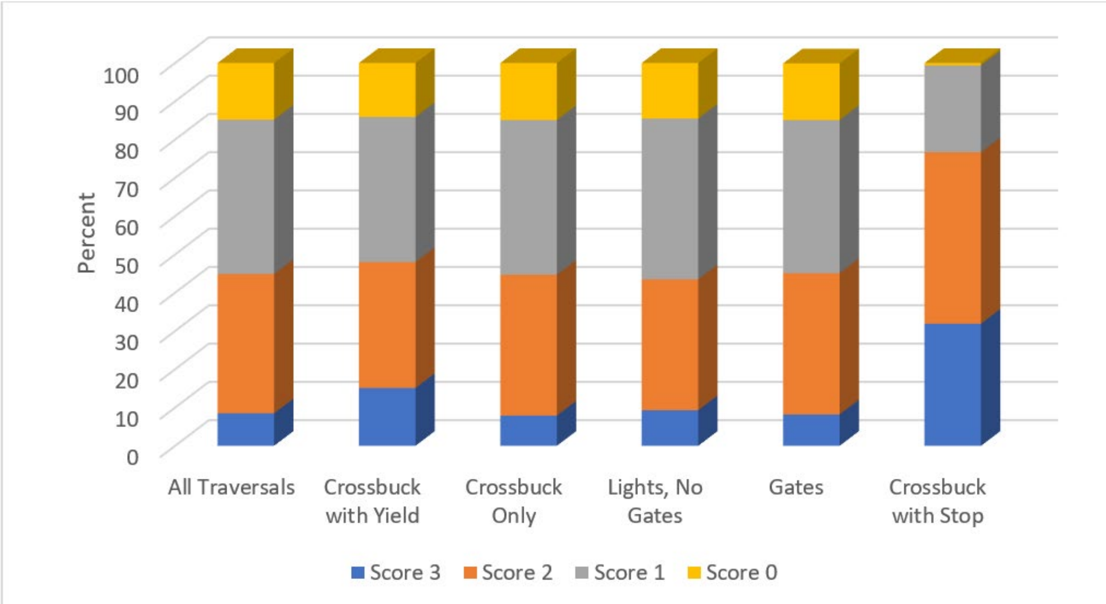


Figure 28. Perecent of Traversals with the Given Total Score, by TCD

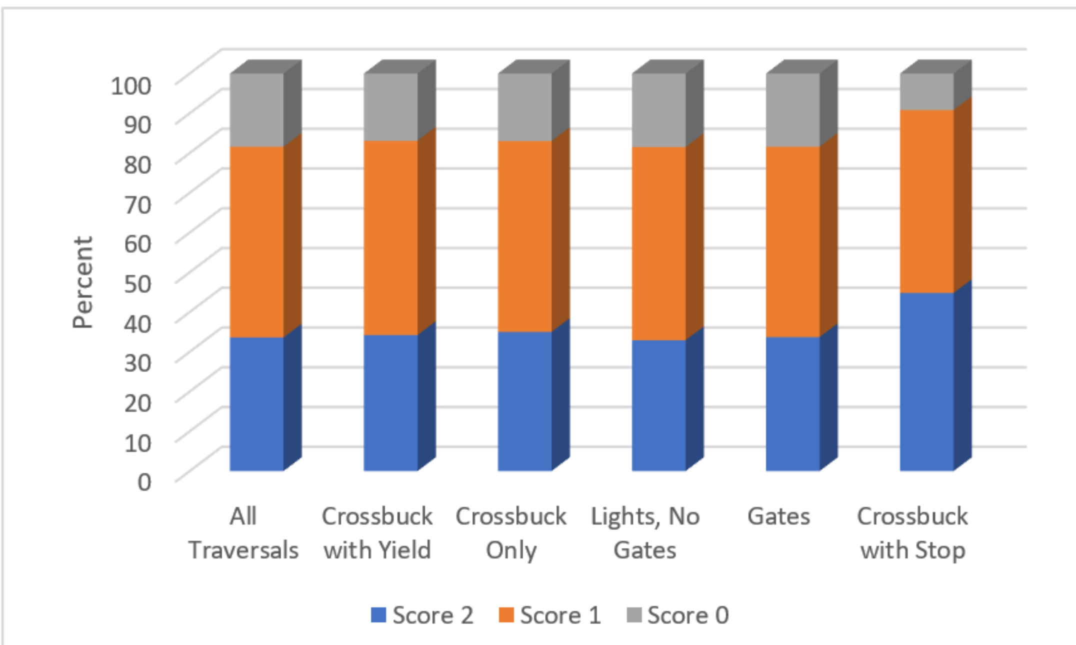


Figure 29. Percentage of Traversals with the Given Scan Score by TCD

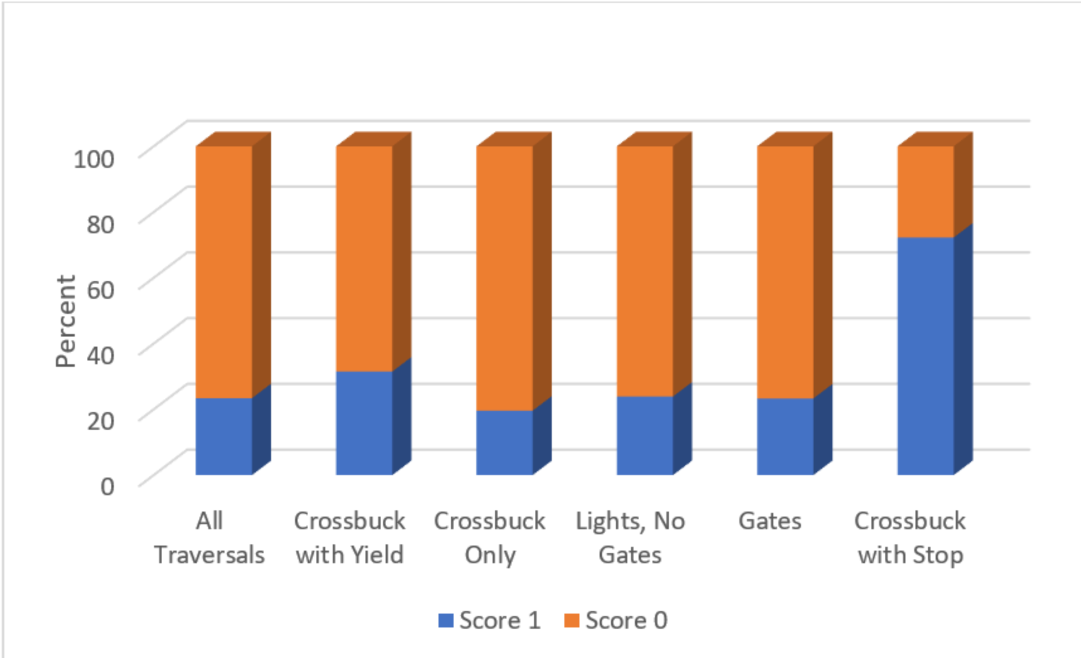


Figure 30. Percentage of Traversals with the Given Speed Score by TCD

Appendix B. Driving Simulator Description

The driving simulator runs the NADS MiniSim version 2.2 software. The hardware (Figure 31) includes a single computer, running Microsoft Windows 7 Pro on an Intel Core i7 processor, 3.07 GHz and 12 GB of RAM, relaying sound through a 2.1 audio system. Three Panasonic TH42PH2014 42-inch plasma displays with a 1280x800 resolution each allow for a 130-degree field-of-view in front of the seated participant. The center monitor is 28 inches from the center of the steering wheel, and the left and right monitors are 37 inches from the center of the steering wheel. The MiniSim also includes a real steering wheel, adjustable car seat, gear-shift, gas and brake pedals, as well as a Toshiba WXGA TFT LCD monitor with a 1280x800 resolution to display the speedometer, etc. Environmental sound effects are also played through two embedded speakers. These sounds included engine noise, brake screech, turn indicators, collisions, auditory alerts, etc. In the present experiment, all participants experienced the same pre-defined route and properties for the driving task.

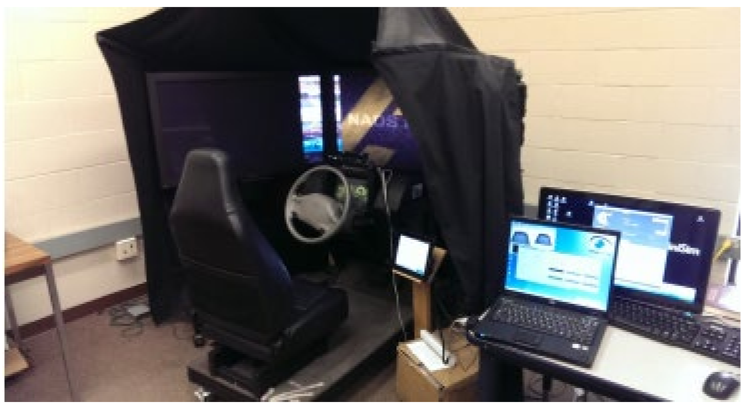


Figure 31. NADS MiniSim Driving Simulator