Transportation Research Part F: Psychology and Behaviour 87 (2022) 313–326

Contents lists available at ScienceDirect

Transportation Research Part F: Psychology and Behaviour

journal homepage: www.elsevier.com/locate/trf

Novel use of a virtual driving assessment to classify driver skill at the time of licensure

Elizabeth A. Walshe [a,*], Michael R. Elliott [b], Daniel Romer [c], Shukai Cheng [a], Allison E. Curry [a,c], Tom Seacrist [a], Natalie Oppenheimer [a], Abraham J. Wyner [c], David Grethlein [d,e], Alexander K. Gonzalez [a], Flaura K. Winston [a,c]

[a] *Children's Hospital of Philadelphia, Philadelphia, PA, USA*
[b] *University of Michigan School of Public Health, MI, USA*
[c] *University of Pennsylvania, Philadelphia, PA, USA*
[d] *Diagnostic Driving, Inc., Philadelphia, PA, USA*
[e] *Computer Science Department, Drexel University, Philadelphia, PA, USA*

A R T I C L E I N F O

Keywords:
New Driver Skills
License Examination
Virtual Driving Assessment
Novice Drivers
Driver Safety

A B S T R A C T

Motor vehicle crash rates are highest immediately after licensure, and driver error is one of the leading causes. Yet, few studies have quantified driving skills at the time of licensure, making it difficult to identify at-risk drivers *before* independent driving. Using data from a virtual driving assessment implemented into the licensing workflow in Ohio, this study presents the first population-level study classifying degree of skill at the time of licensure and validating these against a measure of on-road performance: license exam outcomes. Principal component and cluster analysis of 33,249 virtual driving assessments identified 20 Skill Clusters that were then grouped into 4 major summary "Driving Classes"; i) *No Issues* (i.e. careful and skilled drivers)*; ii) Minor Issues* (i.e. an average new driver with minor vehicle control skill deficits)*; iii) Major Issues* (i.e. drivers with more control issues and who take more risks)*;* and iv) *Major Issues with Aggression (*i.e. drivers with even more control issues and more reckless and risk-taking behavior). Category labels were determined based on patterns of VDA skill deficits alone (i.e. agnostic of the license examination outcome). These Skill Clusters and Driving Classes had different distributions by sex and age, reflecting age-related licensing policies (i.e. those under 18 and subject to GDL and driver education and training), and were differentially associated with subsequent performance on the on-road licensing examination (showing criterion validity). The *No Issues* and *Minor Issues* classes had lower than average odds of failing, and the other two more problematic Driving Classes had higher odds of failing. Thus, this study showed that license applicants can be classified based on their driving skills at the time of licensure. Future studies will validate these Skill Cluster classes in relation to their prediction of post-licensure crash outcomes.

Abbreviations: VDA, Virtual Driving Assessment; GLD, Graduated Driver Licensing.

* Corresponding author at: The Children's Hospital of Philadelphia (CHOP), Center for Injury Research and Prevention (CIRP), 734 Schuylkill Ave, Floor 13, Philadelphia, PA 19146, USA.

E-mail address: walshee@chop.edu (E.A. Walshe).

https://doi.org/10.1016/j.trf.2022.04.009
Received 21 December 2021; Received in revised form 25 February 2022; Accepted 17 April 2022
Available online 29 April 2022
1369-8478/© 2022 Elsevier Ltd. All rights reserved.

1. Background

Motor vehicle crashes –the leading cause of death and injury among US adolescents (Centers for Disease Control and Prevention, 2021; Insurance Institute for Highway Safety, 2021) – are one of the most significant, yet preventable, public health problems. Large population-level epidemiological studies have shown that average crash rates are highest immediately post-licensure and then decline over the first 1 to 2 years of licensed driving (Curry, Metzger, Williams, & Tefft, 2017; Curry, Pfeiffer, Durbin, & Elliott, 2015; Tefft, 2017). Graduated Driver Licensing (GDL) policies have reduced adolescent driver crash incidence principally by delaying licensure to older ages and allowing teens to gain experience under lower-risk conditions, rather than by ensuring adequate skill at the time of licensure (Watson-Brown et al., 2021; Williams, 2017). Despite the increased crash risk for novice drivers, not all new drivers crash, which indicates variable risk across individuals. Identifying those most at risk of crashing allow us to implement precision interventions. (Winston et al., 2016).

With some notable exceptions in California (Chapman et al., 2014) and Ohio, (Walshe et al., 2022) work to date shows that the youngest novice drivers have higher crash risk. (Curry et al., 2017; Tefft, 2017) In New Jersey, Curry et al. quantified the relationship between age at licensure and sex on crash outcomes (Curry et al., 2017; Curry, Hafetz, Kallan, Winston, & Durbin, 2011; Curry et al., 2015), finding that crash rates in the first month of licensure were 50% higher for 17 versus 20 year old (Curry et al., 2015). The authors further found that crash rates for males were higher than for females in the first month of licensure. Furthermore, crash types have been shown to vary by age, sex and experience (Bingham & Ehsani, 2012; Swedler et al., 2012). However, Ohio and California data show a different trend whereby the youngest drivers were less likely to crash, and in Ohio, there were no sex differences in crash rates post-licensure. Of note, both California and Ohio mandate driver education including professional behind-the-wheel training for license applicants under age 18, in addition to GDL restrictions. Furthermore, naturalistic studies of on-road driving using in-vehicle monitoring and lab-based driving simulation studies have shown that one critical reason for over 93% of young driver serious crashes is driver error (Curry et al., 2011; Khattak, Ahmad, Wali, & Dumbaugh, 2021; Seacrist et al., 2021).

Previous attempts have been made to identify young drivers who might be at higher risk for crashes using a variety of data sources (naturalistic, self-report and simulator studies) (Elander & French, 2014; Hooft van Huysduynen, Terken, & Eggen, 2018; Ishibashi et al., 2007). For example, drivers have previously been classified as careful, aggressive/risky, or anxious based on self-reported driving behaviors (and more so habits and risk-taking choices than skills) (Taubman-Ben-Ari et al., 2004). Others have analyzed naturalistic driving behavior in order to categorize driving and/or identify at-risk driver (Feng et al., 2017; Gershon, Ehsani, & Zhu, 2018; Khattak et al., 2021). However, these studies require capture of actual driving behavior and, therefore, are lagging indicators of risky driving; given that crash risk is highest in the month after licensure. A safer approach would be to categorize and identify at-risk drivers based on their skills prior to licensure, before they begin independent driving and enter the period of highest risk (Curry et al., 2015). However, few have focused on identifying specific skill deficits that predict future on-road driving outcomes. Driving simulation provides an opportunity to safely assess skill deficits under risky conditions and thus to potentially identify drivers at risk for on-road collisions (McDonald et al., 2012; McDonald, Kandadai, & Loeb, 2015). However, driving simulation studies, to date, are typically laboratory-based with convenient and small samples that may limit generalizability and categorization of drivers according to future behavior (Mcdonald et al., 2013). Categorizing driving skills (and risk) at the time of licensure through simulated assessment, if given to a large population, has the potential to identify at-risk drivers *before* they begin independent driving.

We collaborated with the State of Ohio where a Virtual Driving Assessment (VDA) has been implemented in the licensing workflow so that license applicants must complete the VDA immediately before their Road Safety Examination (RSE: i.e. the license examination) (Walshe et al., 2020, 2008). This allowed for the first population-level study classifying degree of virtual driving skill at the time of licensure and validating these against a measure of on-road performance: license exam outcomes. The VDA evolved from a validated laboratory-based simulated driving assessment (McDonald et al., 2015) to a shorter and more scalable assessment that can be delivered in the field. The VDA provides ecologically-relevant scenarios based on the most common serious crash scenarios (McDonald et al., 2012; National Highway Traffic Safety Administration, 2018) to identify deficits in operational skills (e.g., basic car control and lane position) and tactical skills (e.g., following distance, gap selection, and hazard awareness and avoidance) (Michon, 1985; Walshe, Romer, Kandadai, & Winston, 2020) (for more details on the scientific foundation and development of the VDA, see (McDonald et al., 2012; McDonald et al., 2015; Walshe et al., 2020). Our prior work in Ohio used the user-generated time-series data from the VDA (vehicle control inputs and other attributes) and a novel application of time-series data clustering methods to automatically classify applicants likely fail the RSE (Grethlein et al., 2020) The resultant VDA score is highly accurate at classifying those who go onto fail the RSE. However, the current study takes a different approach using the derived driver performance variables to identify distinct classes of virtual skill deficits at the time of licensure that can inform targeted interventions, and which show validity in their relationship to (i) subsequent performance on the on-road examination for licensure as determined by Ohio's driving evaluators (criterion validity), as well as (ii) known risk factors for crashes, including age and sex (construct validity).

We hypothesize that there will be some classes representing (i) careful or skilled drivers, (ii) anxious or uncontrolled/unskilled drivers, and (iii) risk-taking drivers, and that these classes will be differentially associated with licensing outcomes and other crash-risk factors. Following prior work in Ohio showing that the youngest drivers under age 18 years, who are mandated to complete driver education including behind-the-wheel driving instruction, had both lower fail rates on the licensing examination and lower crash rates in the early months post-licensure, we expect drivers under age 18 to be more likely to demonstrate skilled and safe performance than those age 18 (Walshe et al., 2022). Conversely, we expect 18- and 19-year old drivers to be more likely to be unskilled (poor vehicle control). Lastly, we hypothesize that the male drivers will be more likely to belong to the risky driving class that demonstrates more speeding and tailgating. (Bingham & Ehsani, 2012; Scott-Parker, Hyde, Watson, & King, 2013).

2. Materials and methods

2.1. Dataset

A data operations team at the Children's Hospital of Philadelphia (CHOP) prepared a de-identified analytic dataset in accordance with data privacy agreements between the State of Ohio and CHOP. Thus, this study was considered exempt from IRB oversight by CHOP. Our analytical sample consists of 33,249 VDA tests taken by 32,836 individuals immediately before the RSE, between July 2017

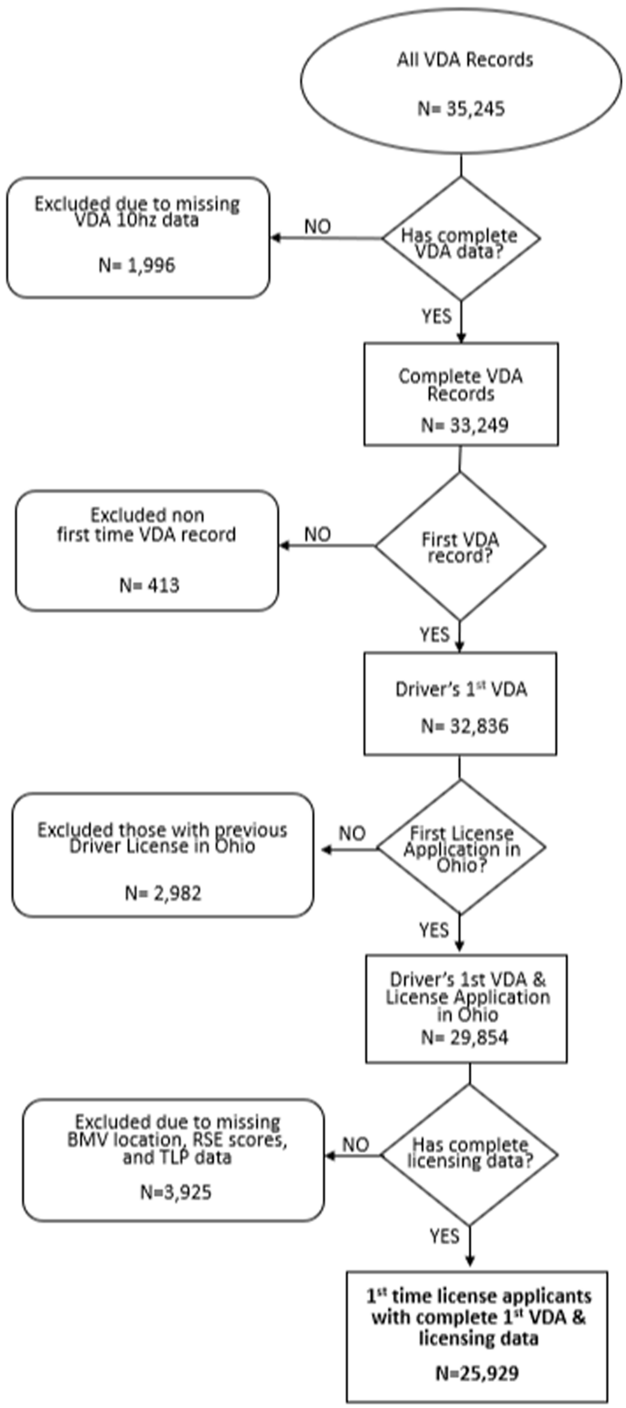

Fig. 1. Sample Derivation Flow Diagram.

and December 2019. We noted that 390 individuals took the VDA twice, 7 took it three times, and 3 took it four times. These VDA tests were taken at five licensing sites in Ohio (three in the urban Columbus area, and one each in suburban Springfield and rural Zanesville). Ohio licensing database records were exact matched to VDA records using a unique driver ID (the temporary instruction permit identification card number), that was then de-identified through an Honest Broker at CHOP. As part of this process, all dates were converted to age in days from the birthdate (no dates were in the final analytical data set); this included age at start of permit, age at first RSE, and age at issuance of license. The driver license addresses were geo-coded into Federal Information Processing (FIPS) codes using ESRI geocoder, and FIPS codes were subsequently matched to American Community Survey (2015) (US Census Bureau. American Community Survey) data to obtain tract-level sociodemographic variables. Some drivers in our sample had missing sociodemographic information because not all addresses could be geo-coded (0.7%). See Fig. 1 for sample derivation.

In order to reduce the large number of available metrics from the VDA into distinct classes of skill-deficits, the first analysis of this study uses principal component and Gaussian mixture model clustering methods on all VDA tests taken from July 2017 through December 2019, including repeat attempts (i.e., no tests were excluded; n = 33,249). The second analysis was to validate the Driving Classes against on-road performance on the RSE. For this analysis, only applicant's first attempt of the RSE and VDA were included (i.e., repeat VDA attempts were excluded). The 2,982 drivers who had previously obtained a license in Ohio were dropped from this analysis. We further restricted our sample by dropping 3,925 applicants who lacked information on their testing location, license examination outcomes, census tract location and learner permit dates, for a total of 25,929 subjects. This is the final analytical sample. Of these 25,929, there were 112 applicants missing median household income data, which appeared to reflective of predominantly single occupant accommodation tract in University areas where household income is not collected because of group quarter (dormitory) residences. To retain these individuals, census tract individual level income data were used in place of the missing median household income values.

2.1.1. Ohio licensing database

This investigation used the VDA test data collected in Ohio (maintained by Diagnostic Driving Inc., Philadelphia, PA) matched to the Ohio state-wide licensing database (maintained by the Ohio Department of Public Safety). The licensing database contains detailed information on each driver's interactions with the Ohio Bureau of Motor Vehicles, including driver demographics (date of birth, gender, address), a driving school completion date, and licensing history (permit dates, RSE dates, license issue dates, and license status). Of note, the gender variable only includes "male", "female" and "unknown" responses (0.04%) so hereafter is referred to as sex. In addition, the database includes outcome variables (pass, fail) for the two RSE subtests: the maneuverability subtest and the driving skill subtest. An overall RSE fail outcome was determined by failure on either or both of the RSE subtests, and this binary variable was

Table 1
Virtual Driving Assessment System variables, definitions, and skill domain.

VDA Skill Domain	VDA Variables	Definition (unit)
Speed-related	Max Speed	Maximum speed value (mph)
	Mean Speed	Mean speed value (mph)
	SD Speed	Standard deviation of speed (mph)
	Max Throttle	Maximum accelerator depression (%)
	Mean Throttle	Mean accelerator depression (%)
	SD Throttle	Standard deviation of accelerator depression (%)
	Drive Duration	Time to complete assessment drive (seconds)
Throttle Control	Max Acceleration	Maximum vehicle acceleration (mph/s)
	Mean Acceleration	Mean vehicle acceleration (mph/s)
	SD Acceleration	Standard deviation of vehicle acceleration (mph/s)
	Max Jerk	Maximum vehicle jerk (mph/s^2)
	Mean Jerk	Mean vehicle jerk (mph/s^2)
	SD Jerk	Standard deviation of vehicle jerk (mph/s^2)
Braking Control	Max Brake	Maximum brake depression (%)
	Mean Brake	Mean brake depression (%)
	SD Brake	Standard deviation of brake depression (%)
Lane Position	Max Heading-Error	Maximum angle between vehicle's heading vector and road-following vector (degrees [0,180])
	Mean Heading-Error	Mean angle between vehicle's heading vector and road-following vector (degrees [0,180])
	SD Heading-Error	Standard deviation of angle between vehicle's heading vector and road-following vector (degrees [0,180])
	Max Lane Deviation	Maximum vehicle lateral displacement from the center of lane (meters)
	Mean Lane Deviation	Mean vehicle lateral displacement from the center of lane (meters)
	SD Lane Deviation	Standard deviation of vehicle lateral displacement from the center of lane (meters)
	Max Road-Center Deviation	Maximum distance from vehicle to center of the road (meters)
	Mean Road-Center Deviation	Mean distance from vehicle to center of the road (meters)
	SD Road-Center Deviation	Standard deviation of distance from vehicle to center of the road (meters)
Route Following	Off-Route	Driving off-road or off-route (count of incidences)
Car Following	Time to Collision < 3 s	Time spent driving at < 3 sec to crash (seconds)
	Miles Driven TTC < 3 s	Distance driven while < 3 sec to crash (miles)
	Time to Collision 3–5 s	Time spent driving at 5–3 sec to crash (seconds)
	Miles Driven TTC 3–5 s	Distance driven at 5–3 sec to crash (miles)
Rule Following	Failures to Stop	Failures to stop at stop signs/red lights (count)
Crash Avoidance	Crashes	Crashes with vehicle, pedestrian or object (count)

used as the ultimate outcome in this study for validating the driving skill classes.

2.1.2. VDA database

The VDA data were acquired from Diagnostic Driving Inc. who collected it in Ohio via the VDA software referred to as Ready-Assess™, which is implemented in the Unity game engine. In a self-directed workflow, (lasting approximately 15 min), license applicants drove within a typical driving route (from a bank of 10 randomly presented routes) that incorporates common serious crash risk scenarios, (McDonald et al., 2012; National Highway Traffic Safety Administration, 2018) including: rear-end events, intersections, curved roads, merges, and hazard zones. These routes also had varied settings (urban and suburban), physical road features, and other road users (for example, crosswalks and merges, construction zones and vehicles, school buses, and pedestrians). A limited number of real-time response measures of driving performance are tracked during the drive and tabulated at 60 Hz within the Unity simulated environment, and a multivariate, time series file is saved at 10 Hz, from which Diagnostic Driving Inc. derived and provided 69 variables that capture aberrant or hazardous driving performance across the drive (e.g., number of red traffic light runs). These 10 Hz variables were designed to capture negative outcomes (e.g. simulated crashes) as well as the nuances of an individual's driving performance in known domains of driving: motor vehicle operation, speed management, collision avoidance, control of the vehicle, and obeying the rules of the road.. Some of these measures, like standard deviation of lane position (SDLP) and time-to-collision (TTC), were inspired by similar efforts to characterize dangerous modes of driving (Verster & Roth, 2011; Vogel, 2003).

All VDA applicant data was indexed by a unique identifier, that was an exact match in the Ohio licensing database, and was securely transferred to CHOP via a data sharing agreement. The study analysis team reduced the VDA dataset to 32 variables that had sufficient variability based on an initial visual inspection of histograms, and these were included in all further analyses. See Table 1 for a list of these variables, a lay definition, and the driving skill domains they capture.

2.2. Statistical analysis

2.2.1. Deriving VDA Clusters and Driving Classes

We created a set of 20 clusters (hereafter, called "Skill Clusters") based on the VDA output for each applicant as follows. First, we used principal component analysis to create summary scores from the 32 VDA measures outlined above. To accommodate the fact that 10 different testing scenarios were used, each of these variables was standardized by subtracting the mean and dividing by the standard deviation of a given scenario, and these standardized variables were used in the principal component analysis. This resulted in 15 principal components (linear combinations of these 32 variables) that explained 90% of the variance in these variables. These 15 components were in turn fed into a Gaussian mixture model, which assumes that each individual's principal component score $\boldsymbol{X}_i$ is drawn from one of K multivariate normal distributions:

$$P(\boldsymbol{X}_i = \boldsymbol{x}) = \sum_{k=1}^{K} \pi_k N_{15}(\boldsymbol{x}; \boldsymbol{\mu}_k, \Sigma_k)$$

where $N_p(\boldsymbol{x}; \boldsymbol{\mu}, \Sigma) = (2\pi)^{-p/2} |\Sigma|^{-1/2} exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{x}-\boldsymbol{\mu})\right)$ is a p-variate normal distribution with mean $\boldsymbol{\mu}$ and variance–covariance Σ. To choose the value of K, we considered the penalized likelihood Bayesian Information Criterion (BIC), choosing $K = 20$ as the value at which the BIC stabilized. The model was fit using the Mclust function in R.

Each subject can be assigned a posterior probability of being in a given Skill Cluster based on Bayes' Rule:

$$P(K_i = k | \boldsymbol{X}_i = \boldsymbol{x}_i) = \frac{P(\boldsymbol{X}_i = \boldsymbol{x}_i | K_i = k) P(K_i = k)}{P(\boldsymbol{X}_i = \boldsymbol{x}_i)} = \frac{\pi_k N_{15}(\boldsymbol{x}_i; \boldsymbol{\mu}_k, \Sigma_k)}{\sum_{k=1}^{K} \pi_k N_{15}(\boldsymbol{x}_i; \boldsymbol{\mu}_k, \Sigma_k)}$$

and assigned to a Skill Cluster based on the modal value $\max_k P(K_i = k | \boldsymbol{X}_i = \boldsymbol{x}_i)$.

We utilized the authorship team's subject matter expertise to review the 20 Skill Clusters (and associated driving performance variable weights), and qualitatively identified patterns of skill deficits, agnostic of the outcome of the RSE. This group consisted of established traffic safety, young driver and simulated driving experts in the fields of public health, epidemiology, engineering, cognitive neuroscience as well as social and developmental psychology. This group reached consensus on identifying four Driving Classes (i.e. four categorical patterns of Skill Cluster deficits reflecting negative or positive weights that varied in strength), with confidence in assigning 15 of the 20 Skill Clusters to these Driving Classes (with uncertainty for classifying five Skill Clusters). A second principal component analysis of the 20 Skill Clusters was conducted to determine which Skill Clusters were closely related and these patterns largely confirmed the classifications made by the subject matter experts, and informed the assignment of the remaining five Skill Clusters to the Driving Classes (where the SMEs were uncertain).

2.2.2. Predicting RSE outcomes

The second analytical step validated the 20 Skill Clusters and 4 Driving Classes against on-road performance on the RSE. Overall RSE Fail was used as the primary outcome variable. We adjusted for licensing center location and considered age groups (16, 17, 18, 19, 20, 21–24, 25–34, 35–49, 50 +) and sex (male/female) as potential confounders, along with time in learner permit (TLP: for the active permit at the time of license examination attempt), tract-level sociodemographic measures summarized by six principal components from the three following categories of Census variables: income/education; transportation; and urbanicity (see Walshe et al., 2022 for more details). We used logistic regression for fail outcomes. To estimate the odds of passing the exams, we used effect coding, so that each category was compared to the overall average score or odds of failing.

We also considered measures of model fit: area under the curve (AUC) for logistic regression. This is a measure of the ability of the Skill Clusters, licensing center location, age, sex, TLP, and sociodemographic factors to classify the testing outcomes of license applicants. The AUC can be interpreted as the probability that the passing applicants in a random pair of passing and failing applicants will have the higher probability (of ½ if they are tied), and ranges from 0.5 (if there is no relationship between the predicted probabilities and the outcome) to 1.0 (if the model perfectly classifies passing and failing). All analyses are adjusted for the site where the test was taken. To account for potential correlations among drivers at each location, a generalized estimating equation model is used with a working independence matrix. All analyses were completed using SAS V9.4.

Table 2
Sample demographics of first-time license applicants who also took the VDA in Ohio, with breakdown by those who passed and failed the RSE (number and percentage of applicants is presented).

Variable	All Applicants	First Time Applicants	Applicants who Passed RSE	Applicants who Failed RSE
All	**N = 32,836 (100%)**	**N = 25,929 (100%)**	**N = 18,389 (100%)**	**N = 7,540 (100%)**
Age				
16 years	8314 (25.38%)	7717 (29.76%)	5964 (32.43%)	1753 (23.25%)
17 years	2700 (8.24%)	2410 (9.29%)	1686 (9.17%)	724 (9.60%)
18 years	4181 (12.77%)	3627 (13.99%)	2355 (12.81%)	1272 (16.87%)
19 years	2050 (6.26%)	1717 (6.62%)	1051 (5.72%)	666 (8.83%)
20 years	1339 (4.09%)	1090 (4.20%)	642 (3.49%)	448 (5.94%)
21–24 years	3555 (10.85%)	2351 (9.07%)	1442 (7.84%)	909 (12.06%)
25–34 years	6132 (18.72%)	3912 (15.09%)	2806 (15.26%)	1106 (14.67%)
35–49 years	3530 (10.78%)	2425 (9.35%)	1911 (10.39%)	514 (6.82%)
50 + years	951 (2.90%)	680 (2.62%)	532 (2.89%)	148 (1.96%)
Sex Male (%)	17,545 (53.43%)	13,927 (53.71%)	10,175 (55.33%)	3752 (49.76%)
Time in Learner Permit				
< 14 days	7255 (23.89%)	5416 (20.89%)	4258 (23.16%)	1158 (15.36%)
14 days to < 6 months	9790 (32.24%)	8443 (32.56%)	5560 (30.24%)	2883 (38.24%)
6–12 months	13,319 (43.86%)	12,070 (46.55%)	8571 (46.61%)	3499 (46.41%)
Tract-Level Median Household Income				
Low: bottom 10th percentile	5649 (17.41%)	4137 (15.96%)	2965 (16.12%)	1172 (15.54%)
10th – 90th percentile	22,892 (70.57%)	18,092 (69.78%)	12,730 (69.23%)	5362 (71.11%)
High: top 10th percentile	3899 (12.02%)	3700 (14.27%)	2694 (14.65%)	1006 (13.34%)
Tract-Level College Degree Education				
Low: bottom 10th percentile	3142 (9.64%)	2458 (9.48%)	1770 (9.63%)	688 (9.12%)
10th – 90th percentile:	22,114 (67.88%)	17,622 (67.96%)	12,578 (68.40%)	5044 (66.90%)
High: top 10th percentile	7324 (22.48%)	5849 (22.56%)	4041 (21.98%)	1808 (23.98%)
Neighborhood Urbanicity				
Urban	10,305 (31.63%)	7806 (30.11%)	5385 (29.28%)	2421 (32.11%)
Suburban	18,732 (57.50%)	15,119 (58.31%)	10,735 (58.38%)	4384 (58.14%)
Rural	3543 (10.87%)	3004 (11.59%)	2269 (12.34%)	735 (9.75%)
Tract-level Average Number of Vehicles per Household	0.482	0.482	0.482	0.482

Note: Average number of vehicles per household was derived by taking the average number of vehicles per tract, and taking the average of all tracts. There was some minor data missing on age, sex, licensing and SES variables within our sample which is detailed in the Derivation of Sample Fig. 1.

3. Results

3.1. Sample demographics

Table 2 describes the sample demographics for all first-time license applicants who also took the VDA at one of 5 licensing centers (described above) in Ohio, and the sub samples of those who passed and failed the RSE on first attempt. Two thirds of the applicants were under the age of 25 years. Drivers aged 16 years made up 25% of the sample, with fewer applicants aged 21 years and above. This sample is skewed older compared to the entire state-wide license applicant population age distribution (see prior work (Walshe et al., 2022; Walshe et al., 2020). Relative to tract distribution across the whole state, this sample was slightly more highly educated and from

Table 3
Driving Classes and Skill Clusters, comprised of performance on VDA metrics across 8 skill domains.

		Virtual Driving Assessment Skill Domain							
Driving Classes, Skill Clusters	All VDA 32,836 (100%)	Speed Control	Throttle Control	Brake Control	Lane Position	Route Following	Car Following	Rule Following	Crash Avoidance
No Issues	**7841 (23.88%)**	+	+	+	+		+		
1. Good Steering Control	*2757 (8.40%)*				+				
2. Cautious	2757 (8.40%)	+	+	+			+		
3. Good Brake & Steering	*2327* (7.09%)			+	+				
Minor Issues	**11,506** (35.04%)	–	–	–	+				
4. Skilled with Hard Throttle	2759 (8.40%)	–			+				
5. Jerky Braking	2092 (6.37%)	–	–		+				
6. Speeder, Tailgater, Rule Breaker	1930 (5.88%)	–					–	–	
7. Skilled Average	*1880* (5.73%)				+				
8. Below Average Control	1753 (5.34%)			–	–				
9. Quick with Controlled Braking	1594 (4.85%)	–		–	–				
10. Mild Control Issues	*1428* (4.35%)			+	–				
Major Issues	**5449** (16.59%)	+	+	+	–	–	+	–	
11. Skilled Rule Breakers	*2441* (7.43%)		+	+				–	
12. Slow, Poor Control, Rule Breakers	1220 (3.72%)	+	+		–		+	–	
13. Less Control, Rule Breakers	*982* (2.99%)		–	+	–			–	
14. Extremely Slow, Poor Control, Rule Breaker	806 (2.45%)	+	+	+	–	–	+	–	
Major Issues with Aggression	**8040** (24.49%)	–	–	–	–	–	–	–	–
15. Aggressive Tailgaters	1352 (4.12%)		–	–			–	–	
16. Extremely Aggressive, Reckless	1317 (4.01%)	–	–	–	–		–	–	
17. Controlled Jackrabbit	1265 (3.85%)	–	–			–			
18. Less Controlled Jackrabbit	1020 (3.11%)		–		–	–			
19. Risky, Poor Control, Jackrabbit	815 (2.48%)	–	–	–	–	–		–	–
*20. Risky, No Control, Jackrabbit**	341 (1.04%)	–	–		–	–	+	–	–

Note: Sample size is all 1st time VDA assessments. The + and – indicate positive or negative loading on each skill domain (for example, a "+" on speed control indicates good speed control, and a "-" indicates poor speed control). Weight strengths are not indicated here but please see Appendix for more detailed cluster loading information. *The *Risky, No Control, Jackrabbit* Skill Cluster is distinguished from *Risky, Poor Control, Jackrabbit*, by higher negative loadings on throttle control and lane position.

Table 4
Percent of drivers who fail the RSE subtests and overall, with odds ratio of failing by Driving Classes and Skill Clusters, adjusting for licensing center location (“Center”) and when controlling for covariates.

Driving Classes, Skill Clusters	**Maneuverability Subtest**			**Driving Skill Subtest**			**Overall RSE**		
	% of Drivers who Fail	OR Adjusted for Center (CI: 95%)	OR Adjusted for Center & Covariates (CI: 95%)	% of Drivers who Fail	OR Adjusted for Center (CI: 95%)	OR Adjusted for Center & Covariates (CI: 95%)	% of Drivers who Fail	OR Adjusted for Center (CI: 95%)	OR Adjusted for Center & Covariates (CI: 95%)
No Issues	13.9	**0.78**	**0.8**	12.70	**0.63**	**0.64**	23.1	**0.63**	**0.71**
		(0.75, 0.80)	**(0.76, 0.83)**		**(0.61, 0.65)**	**(0.63, 0.66)**		**(0.61, 0.65)**	**(0.69, 0.72)**
1. Good Steering Control	13.1	**0.69**	**0.73**	11.30	**0.51**	**0.53**	21.2	**0.58**	**0.61**
		(0.65, 0.73)	**(0.68, 0.77)**		**(0.48, 0.54)**	**(0.48, 0.58)**		**(0.56, 0.6)**	**(0.57, 0.64)**
2. Cautious	13.0	**0.69**	**0.69**	12.80	**0.61**	**0.61**	22.5	**0.65**	**0.64**
		(0.62, 0.78)	**(0.62, 0.78)**		**(0.55, 0.69)**	**(0.53, 0.69)**		**(0.57, 0.73)**	**(0.58, 0.72)**
3. Good Brake & Steering	15.9	**0.87**	**0.89**	14.30	**0.68**	**0.68**	26.0	**0.77**	**0.77**
		(0.85, 0.91)	**(0.86, 0.92)**		**(0.59, 0.78)**	**(0.61, 0.76)**		**(0.68, 0.87)**	**(0.71, 0.84)**
Minor Issues	15.5	**0.91**	**0.93**	16.80	**0.84**	**0.87**	27.3	**0.84**	**0.89**
		(0.86, 0.96)	**(0.88, 0.98)**		**(0.81, 0.88)**	**(0.84, 0.89)**		**(0.81, 0.88)**	**(0.85, 0.92)**
4. Skilled with Hard Throttle	15.1	**0.85**	**0.91**	15.00	**0.69**	**0.73**	25.8	**0.76**	**0.8**
		(0.79, 0.92)	**(0.85, 0.97)**		**(0.58, 0.82)**	**(0.66, 0.79)**		**(0.66, 0.87)**	**(0.74, 0.87)**
5. Jerky Braking	16.0	0.95	0.99	18.80	0.91	0.98	29.1	0.91	0.97
		(0.88, 1.03)	(0.93, 1.05)		(0.77, 1.07)	(0.86, 1.09)		(0.82, 1.02)	(0.9, 1.05)
6. Speeder, Tailgater, Rule Breaker	18.9	1.09	1.07	20.20	0.98	0.95	32.4	1.03	1.0
		(0.99, 1.20)	(0.98, 1.16)		(0.86, 1.13)	(0.82, 1.11)		(0.95, 1.11)	(0.92, 1.08)
7. Skilled Average	13.0	**0.69**	**0.71**	13.70	**0.65**	**0.67**	23.1	**0.67**	**0.69**
		(0.65, 0.74)	**(0.67, 0.75)**		**(0.59, 0.72)**	**(0.62, 0.74)**		**(0.64, 0.7)**	**(0.65, 0.72)**
8. Below Average Control	15.7	0.9	0.94	15.40	**0.71**	**0.73**	26.4	**0.77**	**0.81**
		(0.80, 1.00)	(0.85, 1.03)		**(0.62, 0.81)**	**(0.65, 0.82)**		**(0.69, 0.86)**	**(0.73, 0.9)**
9. Quick with Controlled Braking	13.9	**0.76**	**0.8**	14.70	**0.68**	**0.69**	24.9	**0.72**	**0.74**
		(0.64, 0.90)	**(0.68, 0.94)**		**(0.61, 0.76)**	**(0.63, 0.75)**		**(0.65, 0.79)**	**(0.69, 0.8)**
10. Mild Control Issues	15.3	0.89	0.85	20.60	1.04	1.01	29.9	0.95	0.92
		(0.75, 1.06)	(0.72, 1.01)		(0.86, 1.26)	(0.82, 1.24)		(0.78, 1.16)	(0.74, 1.13)
Major Issues	19.4	**1.24**	**1.18**	25.30	**1.47**	**1.42**	36.7	**1.47**	**1.34**
		(1.15, 1.34)	**(1.1, 1.27)**		**(1.34, 1.61)**	**(1.3, 1.56)**		**(1.34, 1.61)**	**(1.23, 1.45)**
11. Skilled Rule Breakers	18.4	1.07	1.01	23.20	**1.24**	1.17	34.7	**1.19**	1.11
		(0.89, 1.29)	(0.85, 1.2)		**(1.07, 1.42)**	(0.99, 1.38)		**(1.06, 1.33)**	(0.99, 1.25)
12. Slow, Poor Control, Rule Breakers	18.1	1.14	1.15	26.30	**1.47**	**1.59**	36.1	**1.32**	**1.39**
		(0.96, 1.36)	(0.96, 1.37)		**(1.33, 1.63)**	**(1.46, 1.73)**		**(1.21, 1.6)**	**(1.21, 1.6)**
13. Less Control, Rule Breakers	20.6	**1.32**	**1.24**	24.10	**1.3**	**1.21**	37.5	**1.39**	**1.29**
		(1.15, 1.51)	**(1.1, 1.41)**		**(1.21, 1.4)**	**(1.13, 1.31)**		**(1.34, 1.43)**	**(1.2, 1.39)**
	23.4			33.10			43.7		

(*continued on next page*)

Table 4 (*continued*)

Driving Classes, Skill Clusters	Maneuverability Subtest			Driving Skill Subtest			Overall RSE		
	% of Drivers who Fail	OR Adjusted for Center (CI: 95%)	OR Adjusted for Center & Covariates (CI: 95%)	% of Drivers who Fail	OR Adjusted for Center (CI: 95%)	OR Adjusted for Center & Covariates (CI: 95%)	% of Drivers who Fail	OR Adjusted for Center (CI: 95%)	OR Adjusted for Center & Covariates (CI: 95%)
14. Extremely Slow, Poor Control, Rule Breaker		**1.68 (1.49, 1.91)**	**1.58 (1.39, 1.82)**		**2.05 (1.49, 2.82)**	**2.03 (1.44, 2.86)**		**1.85 (1.43, 2.39)**	**1.81 (1.37, 2.39)**
Major Issues with Aggression	18.3	**1.15 (1.11, 1.19)**	**1.14 (1.11, 1.17)**	23.80	**1.28 (1.17, 1.4)**	**1.26 (1.14, 1.39)**	34.4	**1.28 (1.17, 1.4)**	**1.19 (1.21, 1.26)**
15. Aggressive Tailgaters	14.7	**0.86 (0.76, 0.97)**	**0.88 (0.79, 1)**	16.60	**0.79 (0.65, 0.95)**	**0.79 (0.65, 0.96)**	27.3	**0.84 (0.76, 0.92)**	**0.85 (0.78, 0.94)**
16. Extremely Aggressive, Reckless	19.7	1.19 (0.99, 1.43)	**1.19 (1.01, 1.4)**	25.80	**1.33 (1.17, 1.51)**	**1.3 (1.14, 1.47)**	36.7	**1.25 (1.1, 1.42)**	**1.22 (1.08, 1.39)**
17. Controlled Jackrabbit	15.1	0.84 (0.71, 1.01)	**0.83 (0.69, 0.99)**	21.70	1.08 (0.96, 1.21)	1.04 (0.93, 1.17)	31.2	0.98 (0.85, 1.13)	0.95 (0.81, 1.11)
18. Less Controlled Jackrabbit	20.2	**1.27 (1.14, 1.42)**	**1.24 (1.11, 1.39)**	23.60	**1.21 (1.1, 1.34)**	**1.19 (1.03, 1.37)**	35.0	**1.20 (1.12, 1.28)**	**1.17 (1.06, 1.29)**
19. Risky, Poor Control, Jackrabbit	24.9	**1.74 (1.39, 2.19)**	**1.71 (1.34, 2.18)**	33.60	**1.98 (1.73, 2.27)**	**1.92 (1.71, 2.15)**	45.6	**1.88 (1.73, 2.04)**	**1.84 (1.67, 2.02)**
20. Risky, No Control, Jackrabbit	18.8	1.23 (0.88, 1.72)	1.23 (0.87, 1.73)	31.20	**1.78 (1.07, 2.98)**	**1.78 (1.05, 3)**	39.6	1.49 (0.96, 2.29)	1.48 (0.94, 2.35)

Note: Covariates include age, sex, time in learner permit, and sociodemographic factors. Statistically significant odds ratios at p <.05 are in **bold**.

more urban tracts. Furthermore, relative to those who passed their first RSE, there were slightly more applicants from urban tracts and less from rural tracts that failed their first RSE.

3.2. Driving Skill Clusters

Table 3 describes the 20 Skill Clusters and 4 Driving Classes, each with a succinct descriptive summary name, and their relationships to eight domains of virtual driving skills measured by the VDA. To summarize the output, plus and minus signs indicate positive or negative loading on each skill domain (for example, a "+" on speed control indicates good speed control, and a "-" indicates poor speed control). To see in more detail how the individual VDA metrics loaded onto each Skill Cluster, please see whisker plots in Appendix A: Fig. 1-32 in the supplemental documentation. Most cases were assigned to a Skill Cluster with relatively high probability – 90% had posterior probabilities greater than 0.5 and 42% greater than 0.9 (note that random assignment would yield a probability of 1/20 = 0.05). The following 8 Skill Clusters were particularly well defined: *11. Skilled Rule Breaker*; *13. Less Control, Rule Breakers*; *14. Extremely Slow & Poor Control, Rule Breaker; 16. Extremely Aggressive, Reckless*; *17. Controlled Jackrabbit*; *19. Risky, Poor Control, Jackrabbit*; and *20. Risky, No Control, Jackrabbit.* The least well-defined were Skill Clusters *7. Skilled Average* and *8. Below Average Control.* In general, the *No Issues* Driving Class represents no negative weights on any of the VDA skill domains, the *Minor Issues* Driving Class largely represents some vehicle control and lane position problems in the VDA, and the *Major Issues* Driving Class consists of more lane position issues and route- and rule-following errors. The *Major Issues with Aggression* Driving Class was marked by poor vehicle control and positioning, in combination with rule breaking and crashes in the VDA (in the *Risky* labelled Skill Clusters). All Driving Classes were represented in the sample, with the most common being the *Minor Issues* and *No Issues* Driving Classes (collectively: 59% of VDA records), the least common was the *Major Issues* Driving Class (16%).

3.3. Licensing examination outcomes

Table 4 presents the percentage of drivers who failed the RSE (subtests and overall) and the association between the odds of failing and the 4 *Driving Classes* and 20 *Skill Clusters*, with and without adjusting for licensing center location and covariates. For the most part, there were modest changes in odds ratios (ORs) from the model adjusting for license center to the model including covariates (age, sex, tract-level sociodemographic variables). Overall, the *No Issues* and *Minor Issues* classes had below-average risk of failing the RSE overall, while the *Major Issues* and *Major Issues with Aggression* Driving Classes were more likely to fail the RSE overall. The *Major Issues* Driving Class had the highest odds of failing (OR: 1.34 in the covariate adjusted model). Furthermore, the ORs of failing the RSE overall were distinctly worse from the *No Issues* (OR: 0.71) to *Minor Issues* Classes (OR: 0.89), and from there to the two *Major Issues* (OR: 1.34). Driving Classes. In general, it appears the VDA derived Driving Classes and Skill Clusters were more related to the Driving Skill subtest and overall RSE failure outcome than the Maneuverability subtest. When examining the association between RSE fail outcomes and Driving Classes and Skill Clusters, we find that the AUC measure for the Driving Skill subtest (0.663) is higher than the Maneuverability subtest (0.626), with the overall RSE failure (failure on either or both subtest) as an outcome having an intermediate model fit (0.637). While all Skill Clusters in the *No Issues* class had a similar pattern of significantly lower likelihood of failing the RSE and both subtests, there was some differentiation among the specific Skill Clusters in the other Driving Classes. At the Skill Cluster level, *14. Extremely Slow, Poor Control, Rule Breaker* had the highest odds of failing the RSE (OR: 1.81), followed by the *19. Risky, Poor Control, Jackrabbit* Skill Cluster (OR: 1.84).

3.4. Driving Classes and Skill Clusters by Age

Table 5 shows the estimated modal probability of Skill Cluster membership for each age group with a p-value for differences across age groups. All of the Skill Clusters except for *15. Aggressive Tailgaters* were significantly associated with age group. In general, the *No Issues* class was negatively and monotonically associated with age and this pattern was consistent across the Skill Clusters within this Class. The *Minor Issues* Driving Class was positively associated with age, but there was little variation among age groups under 25, and there was deviation within the specific Skill Clusters. *Major Issues* was also positively associated with age, although the greatest difference was between those under 18 (14%) versus those age 18–55 (18%), versus greater than 55 years (23%). Again, there was some deviation at the specific Skill Cluster level. The *Major Issues with Aggression* Skill Clusters collectively peaked at 19–24 (22% vs. 16–19% for those < 18 and ≥ 25 years), however this pattern wasn't consistent across the specific Skill Clusters, and the *15. Aggressive Tailgaters*

Table 5
Driving Class and Skill Cluster distributions by age for the first VDA for each driver, with p-values for age group differences.

Driving Classes, Skill Clusters	Overall N = 25,929	16 years	17 years	18 years	19 years	20 years	21–24 years	25–34 years	35–49 years	greater than50 years	p value
No Issues	24.8%	31.4%	27.5%	25.6%	22.1%	22.8%	20.8%	19.8%	18.0%	11.6%	**<0.001**
1. Good Steering Control	8.6%	11.3%	9.3%	8.2%	7.0%	6.2%	7.0%	7.0%	7.6%	5.4%	**<0.001**
2. Cautious	8.6%	9.9%	8.6%	8.7%	7.1%	9.7%	7.7%	8.0%	7.3%	4.8%	**<0.001**
3. Good Brake & Steering	7.6%	10.2%	9.5%	8.8%	8.0%	7.0%	6.1%	4.8%	3.0%	1.0%	**<0.001**
Minor Issues	40.8%	38.8%	41.3%	39.4%	38.2%	37.3%	39.4%	43.4%	48.1%	46.3%	**<0.001**
4. Skilled with Hard Throttle	8.7%	11.5%	11.5%	9.3%	7.5%	6.6%	6.2%	6.2%	5.9%	4.4%	**<0.001**
5. Jerky Braking	6.2%	4.4%	5.1%	5.4%	6.0%	6.0%	5.6%	8.1%	9.4%	13.2%	**<0.001**
6. Speeder, Tailgater, Rule Breaker	6.0%	5.9%	6.9%	7.0%	8.3%	7.1%	6.7%	4.6%	4.4%	4.4%	**<0.001**
7. Skilled Average	5.6%	5.7%	5.5%	4.8%	3.9%	4.5%	5.0%	6.4%	7.8%	6.5%	**<0.001**
8. Below Average Control	5.2%	3.5%	3.4%	4.0%	3.5%	4.2%	5.6%	7.9%	9.4%	10.2%	**<0.001**
9. Quick with Controlled Braking	4.8%	4.1%	4.6%	4.3%	4.5%	4.6%	4.6%	5.4%	7.0%	4.7%	**<0.001**
10. Mild Control Issues	4.3%	3.6%	4.4%	4.6%	4.6%	4.3%	5.5%	4.9%	4.1%	2.6%	**<0.001**
Major Issues	16.4%	13.6%	14.6%	17.4%	18.2%	18.2%	18.0%	17.8%	17.5%	23.2%	**<0.001**
11. Skilled Rule Breakers	7.7%	7.6%	8.0%	10.0%	9.3%	9.4%	8.0%	6.2%	5.0%	4.4%	**<0.001**
12. Slow, Poor Control, Rule Breakers	3.5%	2.0%	1.8%	2.3%	2.3%	2.2%	3.3%	5.5%	6.7%	15.3%	**<0.001**
13. Less Control, Rule Breakers	3.1%	2.8%	3.1%	4.0%	4.4%	4.3%	4.0%	2.8%	1.6%	0.7%	**<0.001**
14. Extremely Slow, Poor Control, Rule Breaker	2.1%	1.2%	1.7%	1.0%	2.2%	2.2%	2.7%	3.3%	4.2%	2.8%	**<0.001**
Major Issues with Aggression	18.1%	16.2%	16.6%	17.6%	21.5%	21.6%	22.0%	19.0%	16.4%	18.8%	**<0.001**
15. Aggressive Tailgaters	4.1%	3.9%	4.0%	3.9%	4.3%	4.7%	4.2%	4.4%	3.7%	4.6%	0.7
16. Extremely Aggressive, Reckless	3.9%	3.8%	3.4%	4.7%	5.4%	5.1%	4.7%	3.3%	2.5%	2.8%	**<0.001**
17. Controlled Jackrabbit	3.8%	3.6%	4.3%	4.0%	4.7%	5.1%	4.2%	3.5%	2.7%	3.3%	**<0.001**
18. Less Controlled Jackrabbit	3.0%	2.2%	2.5%	2.3%	3.3%	3.2%	4.2%	3.7%	4.2%	4.8%	**<0.001**
19. Risky, Poor Control, Jackrabbit	2.3%	1.9%	2.0%	2.0%	2.8%	2.4%	3.2%	2.8%	2.4%	2.1%	**<0.001**
20. Risky, No Control, Jackrabbit	0.9%	0.7%	0.4%	0.7%	1.0%	1.2%	1.4%	1.4%	0.8%	1.6%	**<0.001**

Note: Statistically significant p-values are in **bold**, according to a Bonferroni corrected p-value of 0.0014.

was not associated with age group.

Of particular interest are differences between applicants < 18 years, who were subject to driver license policy (required to hold a permit for 6 months and complete driver education including behind-the-wheel training before applying for a license), versus those 18–19 who are exempt from such Ohio license policy. Generally, those under 18 were more likely to be in the *No Issues* and *Minor Issues* Driving Classes and less likely to be in the *Major Issues* and *Major Issues with Aggression* Driving Classes (however there was deviation among the Skill Clusters for all classes except for those in the *No Issues* Driving Class). More specifically, applicants < 18 tended to be in the more positive Skill Clusters than those age 18–19 years, including: *1. Good Steering Control; 2. Cautious; 3. Good Brake & Steering*; *4. Skilled with Hard Throttle* and *7. Skilled Average*. Furthermore, those < 18 years were less likely to be in the 3 of the 4 Skill Clusters in the *Major Issues* Driving Class and *16. Extremely Aggressive, Reckless* Skill Cluster from the *Major Issues with Aggression* Driving Class vs. those 18–19 years.

3.5. Driving Classes and Skill Clusters by Sex

The distribution of Driving Classes and clusters varied by males and females among drivers on their first attempt at the licensing examination: see Table 6. Male drivers were more likely to belong to the *No Issues* and *Minor Issues* Driving Class, while females were more likely to belong to the *Major Issues* Class. There was no statistically significant associations with sex for the *Major Issues with Aggression* Class. However, there was deviation in some of the specific Skill Clusters within each Driving Class. For example, male drivers were more likely to belong to the *15. Aggressive Tailgaters* and *16. Extremely Aggressive, Reckless* skill clusters, whereas females were more likely to belong to the *18. Less Controlled Jackrabbit* cluster of the *Major Issues with Aggression* Class.

4. Discussion

This study presents the first population level study of new driver skills at the time of license examination. Using data from a virtual driving assessment implemented into the licensing workflow in Ohio, the results confirmed that new drivers can be classified into distinct skill clusters, which can then be further collapsed into four major Driving Classes: *No Issues* (i.e. careful and skilled drivers), *Minor Issues* (i.e. an average new driver with minor vehicle control skill deficits), *Major Issues* (i.e. drivers with more control issues and who take more risks), and *Major Issues with Aggression* (i.e. drivers with even more control issues and more reckless and risk-taking behavior). These Driving Classes were associated with subsequent performance on the on-road licensing examination, whereby the *No Issues* and *Minor Issues* classes had lower than average odds of failing, and the other two more problematic Driving Classes had higher odds of failing. Furthermore, the odds ratios only changed modestly when covariates were included in the model, suggesting that the effect of driving skills in the model was robust. This study also noted that a number of drivers who were classified as having *Major Issues* or *Major Issues with Aggression* still went on to pass the Ohio License examination. One way to intervene with these drivers is to deliver a personalized feedback report from the VDA, highlighting where the new driver still has room for improvement (e.g. in car

Table 6
Driving Class and Skill Cluster distribution by sex, including p-values for the null hypothesis of no difference in male and female distributions across Skill Clusters and Classes.

Driving Classes, Skill Cluster	Overall N = 25,929	Male N = 13927, 53.7%	Female N = 12002, 46.3%	p-value
No Issues	24.8%	25.5%	23.9%	**0.011**
1. Good Steering Control	8.6%	9.1%	8.2%	**0.014**
2. Cautious	8.6%	8.3%	8.9%	0.098
3. Good Brake & Steering	7.6%	8.2%	6.9%	**<0.001**
Minor Issues	40.8%	42.4%	39.0%	**<0.001**
4. Skilled with Hard Throttle	8.7%	9.7%	7.6%	**<0.001**
5. Jerky Braking	6.2%	6.6%	5.6%	**0.001**
6. Speeder, Tailgater, Rule Breaker	6.0%	5.9%	6.2%	0.443
7. Skilled Average	5.6%	5.1%	6.3%	**<0.001**
8. Below Average Control	5.2%	5.5%	4.8%	**0.022**
9. Quick with Controlled Braking	4.8%	5.5%	3.9%	**<0.001**
10. Mild Control Issues	4.3%	4.0%	4.6%	**0.031**
Major Issues	16.4%	13.9%	19.2%	**<0.001**
11. Skilled Rule Breakers	7.7%	6.6%	8.9%	**<0.001**
12. Slow, Poor Control, Rule Breakers	3.5%	2.9%	4.2%	**<0.001**
13. Less Control, Rule Breakers	3.1%	3.0%	3.3%	0.138
14. Extremely Slow, Poor Control, Rule Breaker	2.1%	1.5%	2.9%	**<0.001**
Major Issues with Aggression	18.1%	18.2%	17.9%	0.635
15. Aggressive Tailgaters	4.1%	4.5%	3.5%	**<0.001**
16. Extremely Aggressive, Reckless	3.9%	4.4%	3.4%	**<0.001**
17. Controlled Jackrabbit	3.8%	3.6%	4.0%	0.072
18. Less Controlled Jackrabbit	3.0%	2.6%	3.6%	**<0.001**
19. Risky, Poor Control, Jackrabbit	2.3%	2.1%	2.5%	0.076
20. Risky, No Control, Jackrabbit	0.9%	0.9%	0.9%	0.787

Note: significant p-values at < 0.05 are in **bold**.

following distance, speed management, etc.) even if they pass the licensing examination, and recommend continued practice in the early months of licensure.

Age was associated with class membership: the youngest drivers subject to Ohio's comprehensive driver training policy (mandated driver education including behind-the-wheel training for applicants < 18 years and a 6-month permit holding period), were more likely to belong to the *No Issues* and *Minor Issues* Driving Classes (as we hypothesized) which were less likely to fail the licensing examination than the other Classes. This finding is consistent with a prior examination of age-trends in Ohio's state-wide licensing and crash data (Walshe et al., 2022). The results also showed that there were more male than female drivers in two of the aggressive Skill Clusters in the *Major Issues with Aggression* Driving Class (specifically, *15. Aggressive Tailgaters and 16. Extremely Aggressive, Reckless*), supporting our hypothesis and corroborating some prior findings (Bingham & Ehsani, 2012; Scott-Parker et al., 2013). However, there was no overall difference at the aggressive Driving Class level.

Taken together, these results indicate that new drivers can be categorized into distinct Driving Skill Classes that show construct validity (being differentiated by age and sex) and some preliminary evidence for criterion validity (against the on-road examination for licensure). Thus, future users of the VDA can be classified based solely on their performance on the VDA (and without input from subject matter expertise). However, future work is needed to validate these Skill Clusters and Driving Classes against crash outcomes post-licensure and examine driver characteristics beyond skills, age, and sex that may explain differences in driver behavior and performance on the VDA. For example, our prior work has shown that executive function abilities associated with the frontal-lobe and rate of development during adolescence, as well as impulsive personality traits, may explain some of the variance in young driver behavior and crash outcomes (Walshe et al., 2019; Walshe, Ward McIntosh, Romer, & Winston, 2017; Walshe, Winston, & Romer, 2021).

5. Limitations

While most of the categories were well-identified by the VDA data and there were substantial differences in risk of licensing exam failure, the VDA-based Skill Clusters were far from perfect predictors of exam outcomes, with a slight majority of the Skill Clusters classified as most likely to fail still passing, and a some of the Skill Cluster classified as least likely to fail (*Good Steering Control*) still failing. Although a large sample of drivers with a wide age range was used, we must also acknowledge that some of the older drivers may have had past licenses from other states that we could not control for in our analyses. Furthermore, while the VDA captures key tactical and operational driving skills during common and serious crash scenarios, the VDA database provided did not contain visual-scanning metrics that are also important skills that may reveal further individual variability. The VDA has these capabilities and can also deliver performance metrics in crash or hazard specific events, so these will be examined in future studies. Lastly, while the VDA performance was associated with on-road licensing examination outcomes, this is not the gold standard of on-road performance and likely does not accurately reflect crash risk post-licensure. Specifically, a numeric error score cut-off is used to determine failure on the RSE, but this binary outcome does not capture those on the border of failing and passing. Indeed, this study shows that some drivers who pass the RSE still exhibit skill deficits on the VDA. Thus, future work will validate these Skill Clusters against crash outcomes post-licensure.

6. Conclusions

This study supports the construct and criterion validity of the VDA for classifying new drivers according to skill deficits at the time of licensure. These results pave a way for identifying individuals who are at elevated risk for unsafe driving, and informing targeted interventions for improving new driver skills before licensure, immediately before a driver's crash risk peaks. However, future work needs to validate these VDA classifications against crash outcomes post-licensure. In addition, future work needs to examine driver characteristics beyond age and sex that may underlie driving behavior beyond skill deficits (such as cognitive and personality factors).

Funding

Research reported in this publication was supported by the Eunice Kennedy Shriver Institute of Child Health and Human Development of the National Institutes of Health (NIH) under award number 1R21HD099635-01 (NICHD); by the US National Highway Traffic Safety Administration through the Ohio Traffic Safety Office (NHTSA); and by the State of Ohio's Department of Administrative Services, managed by the Ohio Bureau of Motor Vehicles; by the Annenberg Public Policy Center at the university of Pennsylvania; by The Distinguished Chair in the Department of Pediatrics at Children's Hospital of Philadelphia (CHOP); as well as a gift from New Jersey Manufacturers (NJM) Insurance Group to CHOP. Data collection was financed by NHTSA and the Ohio Bureau for Motor vehicles. Data management and analyses were supported jointly by NICHD, Ohio Bureau of Motor Vehicles, NHTSA, and the Annenberg Public Policy Center at the University of Pennsylvania. The preparation of this manuscript was supported jointly by NICHD, NHTSA, the Annenberg Public Policy Center and NJM.

CRediT authorship contribution statement

Elizabeth A. Walshe: Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Michael R. Elliott:** Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – original draft, Writing – review & editing. **Daniel Romer:** Conceptualization, Methodology, Formal analysis, Writing –

original draft, Writing – review & editing, Funding acquisition. **Shukai Cheng:** Methodology, Validation, Formal analysis, Data curation, Software, Visualization, Writing – review & editing. **Allison E. Curry:** Conceptualization, Writing – original draft, Writing – review & editing. **Tom Seacrist:** Validation, Writing – review & editing. **Natalie Oppenheimer:** Data curation, Writing – review & editing, Project administration. **Abraham J. Wyner:** Methodology, Writing – review & editing. **David Grethlein:** Data curation, Writing – review & editing. **Alexander K. Gonzalez:** Data curation, Software, Writing – review & editing. **Flaura K. Winston:** Conceptualization, Methodology, Writing – original draft, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

Author Winston and David Grethlein have an intellectual property and financial interest in Diagnostic Driving, Inc. The Children's Hospital of Philadelphia (CHOP) has an institutional financial interest in Diagnostic Driving, Inc. Diagnostic Driving, Inc., created a virtual driving assessment system that is used in Ohio as an assessment at licensing centers and in driving schools to assess driver skills. Winston serves as the chief scientific advisor of Diagnostic Driving, Inc. This potential conflict of interest is managed under a conflict-of-interest management plan from CHOP and the University of Pennsylvania whereby Winston has no interaction with participants (all field data collection procedures were carried out by Ohio Bureau of Motor Vehicles personnel) and all analyses were reviewed and approved by outside consultants with no intellectual or financial interest (John Bolte, a traffic injury researcher at the Ohio State University, and Nancy Kassam-Adams, a behavioral researcher at CHOP and the University of Pennsylvania). Author Grethlein works as a Data Scientist at Diagnostic Driving Inc. His conflict is managed in the same manor as Dr. Winston: Grethlein has no interaction with participants and was not directly involved in the analysis of this paper, which has been reviewed and approved by outside consultants with no intellectual or financial interest (named above).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.trf.2022.04.009.

References

Centers for Disease Control and Prevention. Teen Drivers: Get the Facts. Centers for Disease Control and Prevention. https://www.cdc.gov/transportationsafety/teen_drivers/teendrivers_factsheet.html. Published 202Accessed June 2, 2021.

Insurance Institute for Highway Safety. Teenagers. Insurance Institute for Highway Safety. https://www.iihs.org/topics/teenagers#by-the-numbers. Published 2021. Accessed June 2, 2021.

Tefft B. Rates of Motor Vehicle Crashes, Injuries, and Deaths in Relation to Driver Age, United States, 2014 – 2015. *AAA Found Traffic Saf*. 2017. https://aaafoundation.org/rates-motor-vehicle-crashes-injuries-deaths-relation-driver-age-united-states-2014-2015/. Accessed July 7, 2021.

Curry, A. E., Pfeiffer, M. R., Durbin, D. R., & Elliott, M. R. (2015). Young driver crash rates by licensing age, driving experience, and license phase. *Accid Anal Prev., 80*, 243–250. https://doi.org/10.1016/j.aap.2015.04.019

Curry, A. E., Metzger, K. B., Williams, A. F., & Tefft, B. C. (2017). Comparison of older and younger novice driver crash rates: Informing the need for extended Graduated Driver Licensing restrictions. *Accid Anal Prev., 108*, 66–73. https://doi.org/10.1016/J.AAP.2017.08.015

Watson-Brown, N., Mills, L., Senserrick, T., Freeman, J., Davey, J., & Scott-Parker, B. (2021). A complex system of learning to drive: The instructor's perspective. *Saf Sci., 136*. https://doi.org/10.1016/j.ssci.2021.105172

Williams, A. F. (2017). Graduated driver licensing (GDL) in the United States in 2016: A literature review and commentary. *J Safety Res., 2017*(63), 29–41. https://doi.org/10.1016/j.jsr.2017.08.010

Winston, F. K., Puzino, K., & Romer, D. (2016). Precision prevention: Time to move beyond universal interventions. *Inj Prev.*. https://doi.org/10.1136/injuryprev-2015-041691

Chapman, E. A., Masten, S. V., & Browning, K. K. (2014). Crash and traffic violation rates before and after licensure for novice California drivers subject to different driver licensing requirements. *J Safety Res., 50*, 125–138. https://doi.org/10.1016/j.jsr.2014.05.005

Curry, A. E., Hafetz, J., Kallan, M. J., Winston, F. K., & Durbin, D. R. (2011). Prevalence of teen driver errors leading to serious motor vehicle crashes. *Accid Anal Prev., 43*(4), 1285–1290. https://doi.org/10.1016/j.aap.2010.10.019

Curry, A. E., Metzger, K. B., Pfeiffer, M. R., Elliott, M. R., Winston, F. K., & Power, T. J. (2017). Motor vehicle crash risk among adolescents and young adults with attention-deficit/hyperactivity disorder. *JAMA Pediatr., 171*(8), 756–763. https://doi.org/10.1001/jamapediatrics.2017.0910

Bingham, C. R., & Ehsani, J. P. (2012). The relative odds of involvement in seven crash configurations by driver age and sex. *J Adolesc Heal., 51*(5), 484–490. https://doi.org/10.1016/j.jadohealth.2012.02.012

Swedler DI, Bowman SM, Baker SP. Gender and Age Differences among Teen Drivers in Fatal Crashes. *Ann Adv Automot Med / Annu Sci Conf*. 2012;56:97. /pmc/articles/PMC3503410/. Accessed November 9, 2021.

Seacrist, T., Maheshwari, J., Sarfare, S., Chingas, G., Thirkill, M., & Loeb, H. S. (2021). In-depth analysis of crash contributing factors and potential ADAS interventions among at-risk drivers using the SHRP 2 naturalistic driving study. *Traffic Inj Prev.*. https://doi.org/10.1080/15389588.2021.1979529

Hooft van Huysduynen, H., Terken, J., & Eggen, B. (2018). The Relation between Self-reported Driving Style and Driving Behaviour. *A Simulator Study, 56*, 245–255. https://doi.org/10.1016/j.trf.2018.04.017

Khattak, A. J., Ahmad, N., Wali, B., & Dumbaugh, E. (2021). A taxonomy of driving errors and violations: Evidence from the naturalistic driving study. *Accid Anal Prev., 151*. https://doi.org/10.1016/j.aap.2020.105873

Elander, J., & French, D. J. (2014). *Behavioral correlates of individual differences in road-traffic crash risk: An examination of methods and findings.*. https://doi.org/10.1037/0033-2909.113.2.279

Ishibashi M, Okuwa M, Doi S, Akamatsu M. Indices for characterizing driving style and their relevance to car following behavior. In: *Proceedings of the SICE Annual Conference*. ; 2007:1132-1137. doi:10.1109/SICE.2007.4421155.

Taubman-Ben-Ari, O., Mikulincer, M., & Gillath, O. (2004). The multidimensional driving style inventory - Scale construct and validation. *Accid Anal Prev., 36*(3), 323–332. https://doi.org/10.1016/S0001-4575(03)00010-1

Gershon, P., Ehsani, J. P., Zhu, C., et al. (2018). Crash Risk and Risky Driving Behavior Among Adolescents During Learner and Independent Driving Periods. *J Adolesc Heal., 63*(5), 568–574. https://doi.org/10.1016/j.jadohealth.2018.04.012

Feng, F., Bao, S., Sayer, J. R., Flannagan, C., Manser, M., & Wunderlich, R. (2017). Can vehicle longitudinal jerk be used to identify aggressive drivers? An examination using naturalistic driving data. *Accid Anal Prev., 104*, 125–136. https://doi.org/10.1016/J.AAP.2017.04.012

McDonald, C. C., Kandadai, V., Loeb, H., et al. (2015). Simulated driving assessment (SDA) for teen drivers: Results from a validation study. *Inj Prev., 21*(3), 145–152. https://doi.org/10.1136/injuryprev-2014-041480

McDonald, C., Tanenbaum, J., Lee, Y. C., Fisher, D., Mayhew, D., & Winston, F. (2012). Using crash data to develop simulator scenarios for assessing novice driver performance. *Transp Res Rec., 2321*, 73–78. https://doi.org/10.3141/2321-10

Mcdonald, C. C., Seacrist, T. S., Lee, Y.-C., Loeb, H., Kandadai, V., & Winston, F. K. (2013). Headway Time and Crashes among Novice Teens and Experienced Adult Drivers in a Simulated Lead Truck Braking Scenario. *Proc Int Driv Symp Hum Factors Driv Assess Train Veh Des., 2005*, 439–445. https://doi.org/10.17077/drivingassessment.1524

Walshe, E. A., Romer, D., Wyner, A. J., et al. (2022). Licensing Examination and Crash Outcomes Postlicensure in Young Drivers. *JAMA Netw Open, 5*(4), Article e228780. https://doi.org/10.1001/jamanetworkopen.2022.8780

Walshe, E. A., Romer, D., Kandadai, V., & Winston, F. K. (2020). A Novel Health-Transportation Partnership Paves The Road For Young Driver Safety Through Virtual Assessment. *Health Aff (Millwood)., 39*(10). https://doi.org/10.1377/hlthaff.2020.00802

Grethlein, D., Winston, F. K., Walshe, E., Tanner, S., Kandadai, V., & Ontanon, S. (2020). Simulator pre-screening of underprepared drivers priot to licensing on-road examination: Clustering of virtual driving test time series data. *J Med Internet Res., 22*(6), Article e13995. https://doi.org/10.2196/13995

National Highway Traffic Safety Administration. (2018). National Motor Vehicle Crash Causation Survey: Report to Congress, 2018. Report No. DOT HS 811-059. Washington, DC: National Highway Traffic Safety Adminsitration.

Michon, J,A. (1985). A critical view of driver behavior models: what do we know, what should we do? In L. Evans, & R. C. Schwing (Eds.), *Human Behavior and Traffic Safety*. Boston, MA: Springer. https://doi.org/10.1007/978-1-4613-2173-6_19.

Scott-Parker, B., Hyde, M. K., Watson, B., & King, M. J. (2013). Speeding by young novice drivers: What can personal characteristics and psychosocial theory add to our understanding? *Accid Anal Prev., 50*, 242–250. https://doi.org/10.1016/j.aap.2012.04.010

US Census Bureau. American Community Survey. Accessed December 8, 2020. https://www.census.gov/programs-surveys/acs.

Verster, J. C., & Roth, T. (2011). Standard operation procedures for conducting the on-the-road driving test, and measurement of the standard deviation of lateral position (SDLP). *Int J Gen Med., 4*, 359. https://doi.org/10.2147/IJGM.S19639

Vogel, K. (2003). A comparison of headway and time to collision as safety indicators. *Accid Anal Prev., 35*(3), 427–433. https://doi.org/10.1016/S0001-4575(02)00022-2

Walshe, E. A., Ward McIntosh, C. M., Romer, D., & Winston, F. K. (2017). Executive function capacities, negative driving behavior and crashes in young drivers. *Int J Environ Res Public Health., 14*(11). https://doi.org/10.3390/ijerph14111314

Walshe, E. A., Winston, F. K., Betancourt, L. M., Khurana, A., Arena, K., & Romer, D. (2019). Working Memory Development and Motor Vehicle Crashes in Young Drivers. *JAMA Netw open., 2*(9), Article e1911421. https://doi.org/10.1001/jamanetworkopen.2019.11421

Walshe, E. A., Winston, F. K., & Romer, D. (2021). Rethinking Cell Phone Use While Driving: Isolated Risk Behavior or a Pattern of Risk-Taking Associated with Impulsivity in Young Drivers? *Int J Environ Res Public Health., 18*(11). https://doi.org/10.3390/ijerph18115640