Driving examiners' views on data-driven assessment of test candidates
An interview study

Driessen, Tom; Picco, Angèle; Dodou, Dimitra; de Waard, Dick; de Winter, Joost

DOI
10.1016/j.trf.2021.09.021

Publication date
2021

Document Version
Final published version

Published in
Transportation Research Part F: Traffic Psychology and Behaviour

Transportation Research Part F: Psychology and Behaviour 83 (2021) 60–79

Contents lists available at ScienceDirect

Transportation Research Part F: Psychology and Behaviour

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

Driving examiners' views on data-driven assessment of test candidates: An interview study

Tom Driessen[a,1], Angèle Picco[b,1], Dimitra Dodou[a], Dick de Waard[b], Joost de Winter[a,*]

[a] *Faculty of Mechanical, Maritime and Materials Engineering, Delft University of Technology, the Netherlands*
[b] *Faculty of Behavioural and Social Sciences, University of Groningen, the Netherlands*

ARTICLE INFO

Keywords:
Semi-structured interview
Driving test
Driving license
Data
Driver assessment
Decision aid
Driving simulator

ABSTRACT

Vehicles are increasingly equipped with sensors that capture the state of the driver, the vehicle, and the environment. These developments are relevant to formal driver testing, but little is known about the extent to which driving examiners would support the use of sensor data in their job. This semi-structured interview study examined the opinions of 37 driving examiners about data-driven assessment of test candidates. The results showed that the examiners were supportive of using data to explain their pass/fail verdict to the candidate. According to the examiners, data in an easily accessible form such as graphs of eye movements, headway, speed, or braking behavior, and color-coded scores, supplemented with camera images, would allow them to eliminate doubt or help them convince disagreeing test-takers. The examiners were skeptical about higher levels of decision support, noting that forming an overall picture of the candidate's abilities requires integrating multiple context-dependent sources of information. The interviews yielded other possible applications of data collection and sharing, such as selecting optimal routes, improving standardization, and training and pre-selecting candidates before they are allowed to take the driving test. Finally, the interviews focused on an increasingly viable form of data collection: simulator-based driver testing. This yielded a divided picture, with about half of the examiners being positive and half negative about using simulators in driver testing. In conclusion, this study has provided important insights regarding the use of data as an explanation aid for examiners. Future research should consider the views of test candidates and experimentally evaluate different forms of data-driven support in the driving test.

1. Introduction

The last decade has seen a vast amount of research on automated driving, spanning areas such as sensor systems (Marti et al., 2019; Schoettle, 2017), computer vision (Ranft & Stiller, 2016; Rangesh & Trivedi, 2019), path planning (González, Pérez, Milanés, & Nashashibi, 2015; Marin-Plaza, Hussein, Martin, & Escalera, 2018), and control (Farag, 2020; Lima et al., 2018). At the same time, there is a growing realization that fully automated driving may not be achieved within the next three to five decades (Litman, 2021; Shladover, 2016; Tabone et al., 2021). While there have already been compelling demonstrations of automated driving without human

* Corresponding author.
E-mail address: j.c.f.dewinter@tudelft.nl (J. de Winter).
[1] Joint first authors.

https://doi.org/10.1016/j.trf.2021.09.021
Received 11 August 2021; Received in revised form 11 September 2021; Accepted 27 September 2021
Available online 18 October 2021
1369-8478/

intervention, even the most advanced prototypes to date need occasional human intervention or behave in unexpected manners (Boggs et al., 2020; Goodall, 2021), suggesting that for the coming decades, drivers still need to be trained and licensed.

Although fully automated driving is not within immediate reach, cars are becoming increasingly computerized data-collection machines. Modern cars collect not only data about ego-vehicle state (e.g., speed, heading) and driver input (throttle, brake, and steering) but also data about the environment (presence of other road users and their speed, lane boundaries, infrastructure) via cameras, lidar, radar, or ultrasonic sensors as part of advanced driver assistance systems (ADAS). In addition, data from the traffic environment are now also often collected via nomadic devices such as MobilEye (Chen et al., 2017) and via dashcams and smartphones (Ahmad et al., 2021; Tummala et al., 2019).

The growing proliferation of computers in cars makes it possible to use these computers to assess driver behavior. The literature shows an increasing number of applications that use forward-facing or driver-facing cameras, sometimes combined with acceleration-based triggers, to detect drowsiness and distraction (Chowdhury et al., 2018; Kashevnik et al., 2019; Lechner et al., 2019; Ramzan et al., 2019; Sikander & Anwar, 2018) and unsafe driving behavior (Hickman & Hanowski, 2011; Mase et al., 2020). Other types of systems rely on in-vehicle data recorders (Shimshoni et al., 2015) or smartphones for driver assessment (e.g., Bergasa et al., 2014; Farah et al., 2014; Musicant & Lotan, 2016; Shanly et al., 2018), and see Michelaraki et al. (2021) for a review on post-trip feedback solutions, including smartphone apps, gamification approaches, and reward schemes. Schöner et al. (2021) proposed a concept where a norm-referenced driver safety score was computed relative to the time headway and time-to-collision distributions of a large highway traffic dataset. Similarly, usage-based insurance, also known as pay-as-you-drive insurance, commonly uses smartphones or dongles to obtain measures of driver risk such as speeding, hard braking, or other kinematic events (Arumugam & Bhargavi, 2019; Händel et al., 2013; Vavouranakis et al., 2017) and reward safe behavior with reduced insurance premiums. Relatedly, in motorsports, data acquisition for assessing driving performance is the norm (Segers, 2014).

With the increasing capabilities of in-vehicle sensors and computers, it may become viable to flag deviant driving behavior automatically. This notion can be traced back to the Generic Intelligent Driver Support (GIDS) project, which proposed a tutor that functioned based on the difference between observed driving behavior and reference driving behavior (Michon, 1993). Adaptive training, intelligent tutoring, and driver profiling have already been available in simulator-based driver training for many years (Boelhouwer et al., 2020; De Winter et al., 2008a; Graesser et al., 2005; Karvonen et al., 2006; Ropelato et al., 2018; Wassink et al., 2006; see Zahabi et al., 2020 for a review on adaptive training in simulators). Today, these tutoring concepts are becoming feasible in actual cars. Fridman et al. (2019) demonstrated a real-time intelligent driving system that supervised a second intelligent driving system: disagreements between the two steering angles were found to be predictive of critical situations (automation-to-manual hand-overs). By extension, it should also be possible to have a similar system detect unexpected manual driving behaviors. In the same vein, researchers have performed on-road studies with a personal assistant for fuel saving (Magaña & Muñoz-Organero, 2015) and with an intelligent driving assistant that used an accident risk map and vehicle telemetry as inputs (Terán et al., 2020). In summary, automated assessment of driving tasks seems within technological reach.

The concept of automated driver testing is not far-fetched, at least when it comes to basic driving skills. The Roads and Transport Authority of Dubai has recently implemented a driving test that uses instrumented cars on a driving range, and where the pass/fail verdict is supposedly provided automatically (Government of Dubai, 2019). In 2018, Microsoft introduced Harnessing AutoMobiles for Safety (HAMS), an automated driver license testing system that relies on a smartphone mounted on the windshield and which produces an assessment without human intervention (Nambi et al., 2018), in an attempt to eliminate bribery of the examiners (Giridharan, 2019; Microsoft, 2021).

The above developments may be of strong interest to driving license organizations, which face long-standing challenges regarding the reliability and validity of their driving tests. One issue is test–retest reliability, estimated by performing the driving test twice with different examiners, and the other is inter-examiner reliability, estimated by using two driving examiners in the car assessing the same candidate. Experiments in these areas are scarce but suggest only low test–retest reliability, presumably because traffic conditions change from test to test (pass/fail congruence of 64% in Baughan & Simpson, 1999, and 63% in Olweus, 1958; the latter as cited in Alger & Sundström, 2013), and logically higher inter-examiner reliability of the same test (72% in Bjørnskau, 2003, and 93% in Alger & Sundström, 2013). The high inter-examiner reliability may reflect high-quality assessment procedures, or as noted by Alger and Sundström, "*One possible explanation for the high examiner agreement in Sweden is that quality in the driving test and consistency of assessment are continuously discussed among examiners.*" (p. 28). But even in the study by Alger and Sundström, which found very high inter-examiner reliability, there were occasional disagreements between the two examiners. For example, there were cases where the interpretation of the severity of the candidate's faults or speed adjustments differed between the examiners, or where there were disagreements about whether the candidate should be penalized for faults on specific tasks or should be assessed more holistically. It is noted that low test reliability may be expected if the driving test admits candidates who are *just* good enough to pass, if there is variability in the testing conditions (traffic, weather, road types), or if the driving test is only short (Baughan & Simpson, 1999; De Winter & Kovácsová, 2016).

Another issue is that of predictive validity. Driving test outcomes are not necessarily good predictors of safe post-license driving, as males have been found to perform better on the driving test than females (Crinson & Grayson, 2005; Mynttinen et al., 2011), even though males are overrepresented in post-license crashes (SWOV, 2016). That said, a recent interview study with 13 driving instructors found that instructors often have a sense about whether the learner driver has a risky attitude, lack of concern for safety, or over-confidence (Watson-Brown et al., 2021). These findings are consistent with a study that found that risky pre-license driving in a simulator can predict self-reported post-license traffic violations 3.5 years later (De Winter, 2013). The above factors suggest that driving instructors and examiners may benefit from driver performance data to complement their verdict in a predictive-valid way, pinpoint driving deficiencies, or contribute to the inter-examiner and interregional calibration of driving norms.

In the Netherlands, prospective drivers follow, on average, 40 h of training at a private driving school before applying for the driving test (Roemer, 2021). The Netherlands uses a test-led model, where the driving test implicitly determines the content of the preceding driver training (Helman et al., 2017). Next to an exam on theoretical knowledge, the driving test, organized by the Dutch Central Office of Driving Certification (CBR), involves 35 min of driving, of which 10 to 15 min using a route navigation system. The candidate is assessed based on seven elements of participation in traffic: driving off, driving on straight and curvy road sections, behavior near and at intersections, merging/exiting, overtaking/moving sideways, behavior near and on special road sections, and special maneuvers. The Dutch driving test has undergone various recent modifications, such as the introduction of hazard perception in the theory test and a self-reflection form to be completed before the on-road test (consistent with the Goals for Driver Education; Hatakka et al., 2002). Supervised driving has been introduced as well since 2011 for drivers who obtained their license between their 17th and 18th birthdays (2todrive, 2021). After their 18th birthday, licensed drivers are allowed to drive independently. In introducing further modifications to the driving test, such as the possible introduction of data-driven assessment, it is important to consider the users of such systems, that is, the examiners. User acceptance is crucial, as was also pointed out by De Waard and Brookhuis (1999) in the context of driver support systems: "*A system may function perfectly in the technical sense, if it is not accepted by the public, it will not be used*" (p. 50). In the context of driver testing, acceptance by examiners is crucial.

In the current study, semi-structured interviews were conducted to examine what driving examiners think about the prospect of data-driven assessment. A broad perspective was taken, where we first asked the examiners how they view the current driving test. Subsequently, the interviews went into depth about specific forms of data-based assessment, starting with simple concepts such as automated recordings of speed infringements. However, we also asked the examiners whether they think their task could be replaced by a computer entirely. The interviews also addressed how and when the assessments should be delivered, e.g., during or after the driving test. Additionally, it was asked whether sharing the driving data with different stakeholders would be a welcome idea, an important topic in the era of computerized cars (De Winter et al., 2019; Pugnetti & Elmer, 2020). Finally, we asked some open-ended questions about whether the examiners think that their organization is open to technological change and whether they think that driving simulators, i.e., tools that allow for accurate data recording, could have a role in driver testing.

2. Methods

2.1. Participants and recruitment

A total of 39 driving examiners were recruited, of whom 2 (P21 & P23) canceled their participation, leaving 37 examiners who participated in this interview study. They all were examiners of the driving license "B", which allows driving cars of up to 3500 kg. Twenty-eight participants were male, and nine were female. The average age of the examiners was 46.8 years ($SD = 9.0$ years), ranging from 31 to 62 years. They had on average 9.0 years of experience as an examiner ($SD = 7.8$, min = 1, max = 29), and 51% ($n = 19$) had worked as a driving instructor before (for an average of 12.8 years, $SD = 7.0$, min = 4, max = 27). The examiners reported performing an average of 37.3 driving tests per week ($SD = 7.7$, min = 0.5, max = 47.5). The examiners were recruited from all 12 provinces of the Netherlands, with at least two examiners per province. They were all employed by the Dutch Central Office of Driving Certification.

An invitation email was sent to 17 driving test managers across the Netherlands, together with a one-page description of the study and its aims. The managers then provided the contact details of examiners willing to participate. Before the interviews, the researchers sent the examiners the informed consent form in Dutch via email. It included the following description of the research aim: "*… to investigate driving examiners' views on the validity of current assessment methods, the possibilities of data-driven assessment, and the type of vehicle data they would like to have to support their judgment*". The research was approved by the Human Research Ethics Committee of the Delft University of Technology (approval number 1418).

2.2. Procedure

The interviews were conducted online via Zoom and Microsoft Teams between the 15th of February and the 1st of March 2021. The interviews were conducted by two first authors. Author 1 conducted the interviews in Dutch, whereas Author 2 interviewed in English because she was not a Dutch speaker. Participants willing and able to be interviewed in English, based on self-evaluation of mastering the English language, were interviewed in English by Author 2, whereas the rest were interviewed in their mother tongue (Dutch) by Author 1. As a result, 11 of the interviews were conducted in English and 26 in Dutch.

Each interview lasted approximately one hour. Consent from the participants was recorded orally before the start of each interview. The video recordings of the interviews were stored separately from the consent recordings, in compliance with the data management plan of the project and privacy regulations. The participants were interviewed during their working hours and did not receive additional compensation.

2.3. Interview structure

The interviews were semi-structured according to an interview guide (see Appendix A). The questions were divided into three parts: (1) examiners' opinions about the current driving test, (2) examiners' opinions about a data-driven driving test, and (3) general questions. The interviewers occasionally asked follow-up questions based on the topics mentioned by the participants.

2.3.1. Examiners' opinions about the current driving test

Examiners were asked what the strengths and weaknesses of the driving test are today. Additionally, they were asked whether the test allows them to assess if a candidate would drive safely later on—to obtain a general idea of the perceived effectiveness of the driving test. Furthermore, it was asked whether the examiners' intuition plays a role in establishing the verdict—to better understand the process examiners go through when evaluating a candidate.

2.3.2. Examiners' opinions about a data-driven driving test

The second part of the interview concerned the possible implementation of data in the driving test. The questions were divided into three topics: (a) examiners' opinions about the use of data, (2) examiners' opinions about the characteristics the data should have, and (3) examiners' views about the future of the driving test.

2.3.2.1. Examiners' opinions about the use of data. Examiners were provided with a short explanation: "*With the development of new technologies in the vehicle, it is possible to monitor drivers' behaviors and obtain data regarding their driving performance*". They were then asked if they thought data of any form could be of help in the driving test. They were encouraged to provide any example they could come up with. After allowing some time for the examiner to come up with examples themselves, the interviewer screen-shared a PowerPoint slide, revealing seven suggestions (Fig. 1) one by one. After an item appeared on the slide, examiners rated its usefulness from 1 (*not useful at all*) to 5 (*very useful*) and were encouraged to explain their rating. The presentation order of the seven suggestions was randomized for each participant.

2.3.2.2. Examiners' opinions about the characteristics the data should have. Next, the interviewers asked when data should be collected and when and how the collected data should be presented to them. Examiners' opinions about sharing the collected data were also asked.

2.3.2.3. Examiners' views about the future of driver testing. The examiners were asked whether a part of their task could be automated. Furthermore, they were asked if they thought that artificial intelligence would one day be able to evaluate a driver completely and what it would take for them to rely on such an automated evaluation. Examiners were also asked whether simulators could play a role in the driving test and how they see the future of driver testing.

2.3.3. Closing questions

The interview was concluded with general questions about the possible views of candidates and the licensing organization about data-driven assessment, whether COVID-19 would affect the future of driver testing, and the use of data in other types of tests, such as for motorcycles or trucks.

2.4. Data processing and analysis

The interviews were fully transcribed in their original language. This was done automatically by submitting the audio channel of the recordings to AmberScript (2021) for the Dutch interviews and by using built-in transcription services of Zoom and Microsoft Teams for the English interviews. Errors in the transcripts were manually corrected by Authors 1 and 2.

2.4.1. Qualitative analysis

While conducting and transcribing the interviews, the interviewers familiarized themselves with the participants' responses, similar to the 'familiarization phase' described by Braun and Clarke (2006). Authors 1 and 2 discussed common responses given by the examiners to decide how to subsequently structure the results.

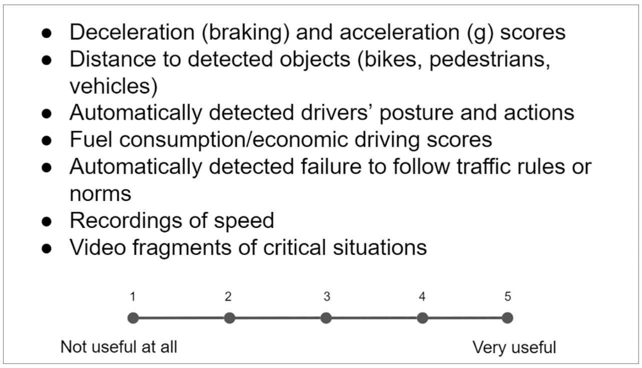

Fig. 1. Slide with seven suggestions of data forms presented.

After further discussions with Authors 3–5, it was decided to retain the structure of the interview guide (Appendix A) for presenting the results. Several questions were omitted (i.e., a question about the differences between driving skill vs. driving style, and the closing questions listed above) since these questions did not appear to yield substantive new insights. Furthermore, the responses to several questions were merged to prevent repetition. For example, responses to questions regarding the recording and presentation of data were grouped in a section entitled: "Introducing data in the driving test: how?" One topic that was not in the interview guide emerged clearly during the interviews. This concerned the examiners' motivations for using data (i.e., as an explanation aid vs. decision aid), addressed in a new section "Introducing data in the driving test: why?"

For each interview topic (e.g., strong points about the current driving test, weak points about the current driving test, future of simulators, etc.), the corresponding quotes were grouped in separate documents. From these documents, quotes were selected to be presented in the narrative of the Results section. The quote selection was done using input from the above-mentioned familiarization phase, based on their potential to explain the topics in a concise manner and by ensuring no over-representation of specific participants. The selection and translation of the presented quotes was done by Authors 1 and 2. Authors 3 and 5 reviewed the selected quotes and corrected the translations using the transcripts.

The above process was applied iteratively, where the results structure was subjected to minor revisions during the quote selection process.

2.4.2. Quantitative analysis

In addition to the qualitative analysis, a quantitative analysis was performed. Firstly, means and standard deviations of the examiners' usefulness ratings of the presented data examples were computed. Secondly, counts were reported for questions that yielded categorizable responses and where counts would illustrate the degree of consensus among participants. More specifically, participant counts were reported of the strong and weak aspects of the exam, examiner-generated data-usage examples, motivations for using data, views about data sharing, and views about the future of driver testing. Responses were tagged (e.g., #strong-aspect.examiner-freedom) in the raw transcripts in Microsoft Word by Authors 1 and 2, a process that was counterchecked by Author 3. Subsequently, the tagged responses were automatically counted using a custom-built Python script (available in Appendix B). This script prevents multiple counts of participants who raised the same item more than once.

3. Results

3.1. Examiners' opinions about the current driving test

On average, the examiners mentioned 1.22 ($SD = 0.47$, min = 1, max = 3) strong aspects and 1.35 ($SD = 0.67$, min = 0, max = 3) weak aspects of the current on-road driving test. Table 1 summarizes the responses. Responses mentioned only once were grouped in the category 'Other'.

3.1.1. Strong aspects

3.1.1.1. Examiner flexibility. Examiners mentioned as a strong aspect that they are allowed some degree of freedom to make judgments and decisions.

> *"… I find it nice that, as an examiner, I am not too bound by all kinds of rules about the assessment. … There is, of course, a framework in which I have to operate, I have to comply with, but I also have my own responsibilities, and I can use my own knowledge and experience to, yes, weigh certain parts. … I also experienced how it was done in the past; it was all with fault codes. Well, I think that is much better*

Table 1
Strengths and weaknesses of the current driving test. *n* is the number of examiners (responses given more than once by the same examiner count as one).

	Response	*n*	%
Strong aspects	Examiner flexibility	14	38
	Human aspect	10	27
	Candidate independence	7	19
	Objective basis	6	16
	Examiner training	2	5
	Other	3	8
Weak aspects	Lack of time	20	54
	Test variability	8	22
	Test is a snapshot	4	11
	Nervous candidates	3	8
	Poor candidate level	3	8
	Test is too strict	2	5
	Other	6	16

now; for example, an item in the past was: 'examiner intervention meant a failed exam'. Well, now, I can think about that for myself. What is my opinion? I am very happy with this." (P7, Translated)

Relatedly, examiners were positive that they are expected to make a holistic assessment of the candidate's ability to drive independently.

"… in the past, we simply had fault scores. Four strikes, or two, three strikes simply meant: failed. Now we have something like an overall image, which means that if the overall image is better than the mistakes made, someone can still pass." (P2, Translated)

"…you are not sitting in a car with an abacus, like: that did not go right, not right, right, right, right, not right, and eventually a number will come out, as it has been in the past. … Now an overall picture emerges. And if someone braked too late, ok, that can happen. If they do that once, compared to the whole ride which they do very well … then I can live with that." (P27, Translated)

Besides having flexibility in the assessment, the examiners pointed out they have the freedom to investigate. Candidates may be guided to different situations based on their performance so far.

"… it is just nice that you currently have the freedom in the exams to adjust your route. The moment I see, for example, [that] you have problems with roundabouts……. Well, then I just take a section with many roundabouts to re-test." (P1, Translated)

"… sometimes people do not drive as they should according to the procedure, but you still feel safe. And you can test it, of course. You take some extra routes, junctions, or anything, and well, I think the strong point is that we have the freedom …" (P10)

3.1.1.2. Human aspect. A strong aspect mentioned by ten examiners is the 'human aspect' of the test. Examiners have to make the candidates comfortable, trying to make them less anxious.

"… it is really important for the examiner; it is our job to get the candidates reassured and make them feel at ease so they can drive how they normally do in driving lessons with their instructor." (P12)

The examiners noted that this comforting facet is new; they mentioned that the guidelines for examiners evolved positively over the past years.

"I think the strong thing about the exams now is that we try to comfort students more now." (P10)

3.1.1.3. Testing candidate independence. Seven examiners mentioned that a strong aspect of the driving test is that it tests for driver independence.

"Well, I think it is strong to the extent that they at least have to show a decent degree of independence." (P29, Translated)

During the exam, candidates are asked to drive independently to a destination specified by the examiner. The candidates use a navigation system for this task. Four of the seven examiners who praised candidate independence mentioned the independent route driving part specifically.

"At the moment, I think one of the strongest aspects [is] that they have to navigate themselves. That was not in it [the driving test] before, and now it is. Because I am still of the opinion that this best approximates how a candidate will eventually behave on the road …" (P15, Translated)

3.1.1.4. Objective standard. Six examiners found the procedure of the driving test a strong aspect. They mentioned that the procedures contribute to objectivity. Some brought up the driving procedure document (CBR, 2020), sometimes referred to as their "Bible" (P39, P4, P36), and noted its positive impact on the driving test.

"It is called the driving procedure. What we expect of the candidate … is all written out in this procedure. All aspects. So, this is the most objective way that we can let candidates take the driving exam. It is very clear: what we expect … and how we judge it." (P4)

3.1.2. Weak aspects

3.1.2.1. Lack of time. Twenty examiners mentioned lack of time as a weak point of the driving test. They indicated that due to high traffic density and decreased speed limits, it has become challenging to (re-)test the desired skills within the 35 min of driving time.

"… when I look at 25 years ago and now, when looking at the traffic intensity, but also the residential areas that are now all 30 km/h zones, then time is sometimes short." (P6, Translated)

"… traffic is so busy that you cannot always test everything. You are constantly thinking: 'okay, … I also have to be back in time for the next test.'" (P3, Translated)

"… I think we need more time. When we want to test well, we need more time to assess [candidates]. … Within the short amount of time, we cannot always do long stretches of highway, stretches outside urban areas …" (P15, Translated)

"[A] shortcoming is that … due to the time, the area in which you drive is restricted, the radius around your place [examination office]." (P29, Translated)

3.1.2.2. Test variability. Eight examiners reported that the variability in testing conditions is a weak point of the driving test. They mentioned that traffic conditions are variable and dependent on the time of day or testing location.

"... we try to do every exam the same, but traffic situations can be completely different. For instance, an exam on Tuesday morning at eight o'clock is completely different from an exam on a Saturday morning at eight." (P18)

"... traffic all around the country is different from place to place. It is a lot harder to do a test over here in the east of the Netherlands." (P8)

Furthermore, individual differences between examiners are a source of variability.

"Of course, because it is a human effort, there will be people [examiners] who make the route more difficult than may actually be necessary." (P1, Translated)

"I think that my colleagues and I can be really all-determining. How I create my atmosphere, or things I could say, well-meant, or maybe not well-meant; with that, you can get somebody to a certain verdict [pass/fail], I think that that may be the weaker aspect of the test." (P2, Translated)

3.1.3. Predicting safe driving

It was asked whether the current driving test allows examiners to predict safe driving later on. Some examiners interpreted this question as to whether the test helps them assess candidates effectively. From this perspective, the answers were generally positive, with a few mentions of the lack of time.

"In general, yes, it does. But well, every now and then, there are times, like I said, that time-wise, [it is] always a bit tight. If you would have more time, the verdict is probably going to still be the same. But sometimes you just need some more time to check; you reevaluate." (P18)

The second interpretation regarded the ability to predict whether a driver will drive safely after having passed the driving test. The examiners generally indicated that they are not well able to predict what will happen in the future.

"Well, in order to give an honest answer to that question, [If] I would be able to know how they drive after the exam. I do not." (P14)

The examiners specifically pointed out that candidates can pretend: they may adopt an appropriate driving style but reveal themselves as aggressive drivers or risk-takers when driving independently.

"... I am sure that the candidates that pass the test because they drive in the way we would like him to drive at that time. But for sure, later on, they will change their attitude in traffic." (P8)

3.1.4. Examiner intuition

The examiners indicated that intuition plays a role but not to the point of deciding on a verdict. Their intuition may, for example, help them assess situations more quickly.

"I do not know if it is intuition or if it is knowledge. Because when you have done more exams, you can recognize sooner where the problem might be or what was good and what is not." (P13)

Furthermore, based on their intuition, or 'gut feeling', examiners may formulate hypotheses to be tested by gathering additional information.

"You sometimes have a particular gut feeling about behaviors or from certain expressions they give. This can mean that you sometimes want to check a certain thing, an intersection or perhaps a highway, by which the behavior you have a gut feeling about, that is, the so-called intuition, may or may not come out positively." (P29, Translated)

Although intuition can influence the route driven, according to the examiners, the final judgment is always based on facts and procedures.

"In principle, we base ourselves on what we see, so the actual facts. We have to judge based on that." (P34, Translated)

"So you cannot let somebody fail because your intuition tells you it is not good enough ... If one fails, you have to tell them facts. And if you do not have any facts, you cannot let them fail." (P39)

3.2. Examiners' opinions about a data-driven driving test

After introducing the basics of driver monitoring, the interviewers asked the participants to express their views about the use of data in the driving test. Some examiners were positive and enthusiastic about the idea:

"I think so; well, I am quite positive." (P7, Translated)

"Yes, absolutely; I totally agree." (P34, Translated)

A small number of examiners based their answers on previous experience with video recordings or driving simulators, when they were still driving instructors:

"That is why we used it [video recordings]. It helps a lot. We saw a big difference between when we were using it and before we were using it." (P39)

"*We had a camera ride ... you can see a lot of things ... it was really helpful because I had some students who were very, well, a bit naughty or stubborn, and they saw themselves, and they were like: 'Whoa! Am I doing that?'*" (P10)

Others were a bit more hesitant or asked for clarification.

"... I think ... my feeling says no ... " (P15, Translated)

"I have been thinking since I received the invitation for this interview what kind of data would that be." (P14)

During the interviews, discussions emerged about different uses of data. The interviews addressed the different purposes of data mentioned during the interviews ('why'), what data examiners would like to use ('what'), and how the specifics of data recording should be arranged, such as delivery, data sharing, and moment of recording ('how').

3.2.1. Introducing data in the driving test: why?

The interviews addressed whether data could help examiners come to their verdict. Two main motivations for the use of data became apparent from the interviews: using the data for explaining the verdict to the candidate (explanation aid) and using the data to support the examiner in arriving at the verdict (decision aid).

3.2.1.1. Explanation aid. The examiners saw merit in the use of data as an explanatory tool. As many as 36 of the 37 examiners mentioned they would want to use data for this purpose. They mentioned encountering candidates who refuse to accept a fail verdict or who even become aggressive or file a complaint. Thus, the examiners would like to have objective data to back up their assessment.

> *"I already make my judgment without all these things, of course…. The only way it would be useful is avoiding the discussion and avoiding the aggressiveness and the angriness, and the one who is going to threaten you or file a complaint. Because they do not have any grounds if you can show them 'look here'…. I would use this after the exam to back up my story."* (P16)
>
> *"If you tell a certain person or candidate that their [following] distance is too short, they will often defend it using the motto 'I think this is sufficient', so to speak. If you can show based on the equipment, how often they have not kept a sufficient distance, for example, that would be an addition."* (P29, Translated)
>
> *"If you assume that an examiner is competent, then you can actually tell, regardless of the data, whether someone is fit to drive … No data is necessary, I would say, because I can just see they have a too short following distance. I can also see whether they are driving too fast. I think that data is very useful to get the candidate to feel: 'yes, that examiner is in fact right; I indeed was not safe there … I indeed did not look properly there. I indeed drove too fast there.'"* (P33, Translated)

3.2.1.2. Decision aid. The interviewers often followed up with questions such as "*And could the data also help you establish the verdict?*" or "*Could the data also help you in making decisions?*", to assess the use of data as a decision aid. Overall, the examiners believed they do their job well already and do not need data to judge a candidate. More than half of the examiners indicated they would not want to use data to reach the pass/fail verdict.

> *"Yes, but not to make the decision. I know how to decide if somebody passes or fails. I do not need data for that… I do not see the possibility, really, yet, to help me make my decision."* (P14)
>
> *"We are at this moment strong enough to come to a verdict."* (P6, Translated)

Several examiners were positive about the use of data as a decision aid, however. They mentioned that it could be useful to obtain extra information, as they cannot pay attention to everything. Data availability could also support examiners' memory and let them review situations through video recordings. It was also argued that data could be used to improve objectivity.

> *"… we cannot follow candidates' eyes during 100% of the test……. We miss some things, I think, sometimes in really important moments. Because we have to be aware and pay attention to the traffic."* (P8)
>
> *"… those technologies are developing rapidly at the moment. … and I think we might be able to be assisted by those systems. Possibly with exam results or as support……. support systems that help us establish a proper result [of the driving test]."* (P24, Translated)
>
> *"… certain things stick with you [in memory]. And why wouldn't it be the case that actually too many negative things stick, and that the positive things do not stick enough, or vice versa. So, I think, if you just set objective data to that, and you can preview it before you present the test outcome … I think you might well be in for some surprises. You might think: 'darn, my own view was different after all.'"* (P11, Translated)

3.2.1.3. Other uses of data. Besides explanation aids and decision aids, the examiners brought up other potential uses of data. Data could also improve the way new drivers are taught, for example, by learning from previous mistakes.

> *"You don't have any discussion because it is all clear; everybody can see it. For the candidate but also for the instructor. And they can, for instance, if it is not good, they can practice with it because they know exactly what happened."* (P18)

A recurring issue the examiners mentioned is that they often encounter students who clearly lack the skills required to pass the driving test. Data about the number of hours of training or the training conditions encountered could be used to preselect candidates for the driving test.

> "*Certain schools often just come with candidates who are far from ready for the driving test.*" (P7, Translated)
>
> "*that you know … that [the candidate] fulfills minimum requirements like 'this many training hours, training at different times of the day, … and this many hours … on the highway.'*" (P38, Translated).

It was further mentioned that data could help create uniform norms and assess the effectiveness of driving test elements, such as special maneuvers. Another suggestion from the examiners was to analyze the routes driven. By integrating this information with traffic intensity data, optimal routes could be generated that bypass congested intersections.

3.2.2. Introducing data in the driving test: what?

3.2.2.1. Data suggestions by examiners. Table 2 lists the examples brought forward by the examiners when asked what data they could use in their work. Furthermore, Fig. 2 shows the ratings of the concepts provided by the interviewers (additional statistics can be found in Appendix C).

Items that received a high rating were typically discussion points between examiners and candidates. Most frequently, the examiners suggested capturing the candidates' eye-gaze behavior.

> *"Another thing is looking, observing, … we can pretty well see in a mirror what somebody is looking at, but it could be supported by data …"* (P19, Translated)

Eye-gaze data could help prevent discussion or misunderstandings.

> *"Gaze behavior, you could look … at what the candidate is looking towards … then [I think] that you can get a lot of misunderstandings out of the way."* (P5, Translated)
> *"We call it 'viewing technique', and that is always a discussion point. When you have a candidate, and you say 'you are not looking, or you are not looking right or enough' … [With] access to data, how they [the candidates] look, … you avoid that discussion because you can show it."* (P12)

The examiners were aware that the eyes pointing somewhere does not imply that the driver perceived the event.

> *"But well, I have seen people looking into a street, looking straight at the car, and who still continue driving."* (P22, Translated)
> *"I cannot see if you really perceive something. I can only tell from the action you make."* (P1, Translated)

The opposite was also noted: things can be perceived even when the candidate does not appear to have moved the eyes towards the object.

> *"I do not see him look anywhere, and still he responds to everything he should respond to. So, well …"* (P9, Translated)

The examiners frequently indicated they would like to have access to recordings of driving speed. They saw value in such recordings, especially if the recordings could be related to the driving context, such as oncoming intersections, curves, or before merging onto the highway.

> *"A simple example is speed. An important item in the driving test is an intersection within the built-up area. These should proceed safely. When you drive too fast there, that's not safe."* (P33, Translated)
> *"I could use this [speed recording] in many cases to explain that someone is approaching too fast, for example, approaching that intersection too fast, approaching too fast on roundabouts."* (P36, Translated)
> *"We are sometimes having a discussion about it [about speed], you know; we have a very nice speedometer we can look at, and yet there is still discussion about how fast someone has driven."* (P19, Translated)

Measuring the distance between the vehicle and other road users/objects was regarded as valuable as well. Distance was seen as relatively easy to measure via sensors.

> "*Take, for example, keeping distance, that two-second rule. You can, of course, measure very well whether … enough distance is kept …*" (P17, Translated)
> "*What average drivers—I am not even talking about novices but also existing drivers—find difficult: … keeping distance. And the modern car of today can show, register, what your current headway is.*" (P29, Translated)

An index of lane-keeping behavior was also regarded as valuable.

> "*Position on the road on straight roads, I could use that, because there are candidates who really zigzag on their way; they just have no feeling of staying in the middle of a lane.*" (P28, Translated)

Cars used for driver training and testing have a double control system. In the case of a hazard, the examiner can (preemptively) press the brakes. Such an intervention can be a source of debate, where data may be of help.

Table 2
Examples of the types of data that could be used, brought forward by examiners. *n* is the number of examiners who brought up the response (responses given more than once by the same examiner count as one).

Response	*n*	%
Gaze behavior	22	59
Recordings of speed	20	54
Distances	10	27
Position on road	8	22
Braking	8	22
Eco-driving	5	14
Video recordings	5	14
Reaction time	3	8
Traffic signs	3	8
Vehicle handling	3	8
Other	12	32

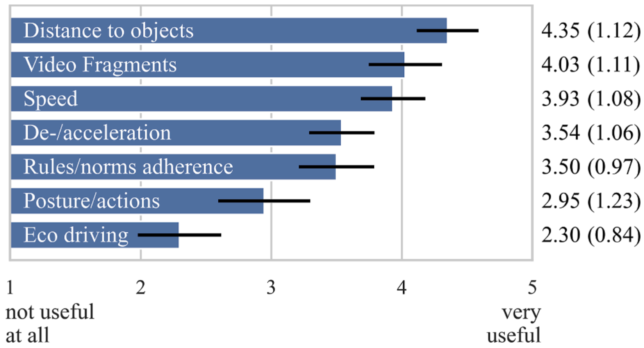

Fig. 2. Means (*SD*) of the ratings of the data examples provided by the interviewers ($n = 37$). Error bars represent 95% bootstrapped confidence intervals computed using Morey's (2008) method for within-subjects designs.

"*We hear from candidates almost daily, like: 'I also braked myself' [after an examiner intervention]. We are generally professional enough to be able to see whether somebody braked or not, or if we have done that [ourselves]. Data can just point that out. Data could simply tell: 'you have or have not braked at that moment.'*" (P19, Translated)

Video recordings were deemed useful for providing feedback to the candidates about traffic violations or aberrant behaviors.

"*Yes, sure, that would be … after the driving test, just like in 'Blik op de weg' [Dutch TV program showing drivers committing traffic violations], where they are able to rewind the violations very quickly, so that you can see: '… in this situation … here you did that, and I did this.'*" (P3, Translated)

"*What you sometimes see when … being stopped by the police, is like 'Please come along, and then we could show you some video footage.'*" (P6, Translated)

The examiners indicated that body position was not of much interest to them.

"*If you can drive a car [sitting] backward better than with your nose forward, then you should do that.*" (P11, Translated)

However, the examiners were relatively enthusiastic about the measurement of head posture to extract eye gaze behavior, which may explain the overall medium rating of posture/actions (Fig. 2).

Even though eco-driving is formally part of the assessment criteria (CBR, 2020), eco-driving was rated the lowest. Examiners reported that, in practice, candidates do not fail the test because of bad eco-driving. Examiners noted that factors causing bad eco-driving are usually bad vehicle handling, such as over-revving or jerky driving, which are factors that are obvious to an examiner without data support.

3.2.3. Introducing data in the driving test: how?

3.2.3.1. When should data be recorded?. Examiners were asked whether data should be collected before, during, or after the driving test. Some examiners answered that the driving test should only consider the ride itself: they would not want to take data from the driving lessons into account, for example.

"No, I do not think that is of any added value for my final test. Each test is on its own." (P5, Translated)

Obtaining additional information about the candidate still appeared interesting to some examiners. They explained that receiving data from previous driving lessons may help reduce the 'snapshot' issue:

"[It is] often that they tell me that it is due to nerves, and I think that if the data show that it is true, that they have done things differently [in the exam] than they have done during the lessons, then it could be helpful in some way …" (P13)

The examiners pointed out there is a conflict of interest with the driving instructor and that they have no control over how the pre-examination data are collected.

"Perhaps I want to say very carefully that I do not trust them [instructors] enough for that. You can manipulate it now; it can be set up. I want it to be objective; it is too important." (P4)

"The difficulty with using the data from the lessons, I cannot see if it is the instructor who says 'oh, you have to slow down' or 'you have to wait' or 'you have to …'. So I do not know if it is from [the candidates] themselves." (P35)

Benefits of data were also identified for the candidates: data about the driving lessons could help them get familiarized with data and facilitate self-reflection.

"I think that if there will be data in the lessons, they will be better prepared. It gives them some self-reflection." (P14)

"*I remember seeing images from a certain driver training course of how I acted in traffic, on a motorcycle … I think that for driving schools, this can be a very useful tool, simply because you have immediate feedback.*" (P36, Translated)

Some benefits of data collection after the driving test were mentioned as well. The examiners pointed out that post-license data collection may help candidates self-reflect and support them in the first years of independent driving.

3.2.3.2. When should data be delivered?. Three options were discussed with the examiners: to receive data before, during, or after the driving test. A clear outcome was that receiving data during the test was judged as impractical. It would only be possible for small amounts of information or information that is very easy to process.

> *"In essence, most of the data we collect ourselves in real-time as a human being. I do not think it is useful to have one more extra stimulation. In that regard, the job is sometimes already busy enough: all your senses are used."* (P36, Translated)
> *"During the exam, we have enough to do. Paying attention to the traffic, paying attention to the candidates, maybe to the instructor—if he or she is allowed to come with us. So, well, maybe about speed, that can be data that you can just read from a screen, maybe, or you can see in a blink of an eye."* (P10)

Real-time feedback based on the data collected was also considered. It could assist examiners in making split-second decisions such as taking the wheel to ensure safety. In that sense, it would be acceptable to receive data while driving.

> *"So this would be helpful if there is, like a sound in the car at the moment, someone, a pedestrian, or a bike is too close, and I would just take the wheel."* (P16)

Some examiners stated it would be useful to adapt the driving test based on the information recorded during the preceding driving lessons. However, there were concerns about not being objective anymore.

> *"I think you could be influenced if you already know someone, what their weak spots are. I would not want to know myself."* (P18)

Receiving the data after the driving test was the most accepted option. In the current driving test, the examiners give their verdict only a few minutes after the driving part. Analyzing the data might require extra time, which means a change in the test setup may be needed.

> *"If data were to be used to make a decision, … then maybe a moment will have to be inserted just after the actual driving where I retreat for a few minutes. I am going to deliberate with myself and with the data I have, to see: 'What do I actually think of that?'"* (P36, Translated)

3.2.3.3. How should data be delivered?. Regarding hardware, the examiners explained that they already use tablets and that these could be very suitable for receiving and presenting data. Receiving information in an auditory way, via headphones, was also mentioned.

According to the examiners, data could take the form of color-coded results or percentages. Lack of time to analyze the raw data was often listed as a reason why examiners preferred processed data.

> *"Because time is always an issue. So it must be easily accessible, easy to read."* (P13)
> *"… if you see those seven elements that we grade, and you would have a stoplight principle, and so red, yellow, and green, where you could actually see back, 'These parts are all green. Only this part is red, and this part is orange. Or yellow' as the law requires. You could still retest that yellow part or that red one."* (P38, Translated)

Some examiners preferred less-processed data, yet still easy to read. Graphs were often given as examples.

> *"… I think something that is at least very simple, that you can see at a glance, and I do not know how that would have to be worked out. But yes, graphs, indeed, often then you can see something pretty quickly without looking at numbers in great detail. I think that might be useful."* (P28, Translated)

However, a few examiners preferred having access to a large amount of information, insisting that it is their task, not the computer's task, to analyze the information.

The examiners further indicated that having access to single variables only is not particularly helpful, as single variables lack the necessary context. The examiners pointed out that this problem could be resolved, in part, by presenting a combination of data types or through additional information (e.g., location, traffic density, video).

> *"I would like to see the combination, with a map, so the driven route…. Combination with data you have, deceleration, distance, so that you get a piece of video."* (P25, Translated)
> *"Whether the candidates … should fail or pass should be our judgment. I think graphics would help. So regarding the G forces or brake and acceleration forces and the video fragments. So I think it is a combination of that."* (P39)

3.2.3.4. With whom should data be shared?. The possibility of data sharing was discussed during the interviews. Sharing with the candidate, the instructor, other examiners, or the testing organization were proposed. The examiners were generally in favor of data sharing, especially with the candidate. Out of the 35 examiners who mentioned the possibility of sharing data with the candidate, 34 were in favor. The main arguments were that the data belongs to the candidate and that data could have educational purposes.

> *"I think anyway that if I am allowed to see it, that the candidate should be able to see it as well, because it is his behavior, his exam …"* (P2, Translated)
> *"Well, of course we all do it for traffic safety, so it makes little sense to only share that [data] with me. Of course [it is] very useful … for the candidate to take note [of the data] and learn from it."* (P31, Translated)

When discussing the possibility of sharing data with other stakeholders, some privacy issues were raised, and sharing was only considered viable if the data were made anonymous.

> *"I would not have a problem with that [sharing], as long as it does not haunt the person themselves…. because it is a snapshot, people are very vulnerable. And then it would not be fair."* (P22, Translated)

The interviewers suggested that, once anonymized, data could be used to improve uniformity across driving tests, for example, by identifying discrepancies between examiners and to train them. This type of data use was generally agreed upon, although some examiners expressed concerns regarding potential misuse and an increasing number of complaints. The examiners also noted that it is hard to compare individual driving tests in the attempt to achieve uniform assessment criteria.

3.3. Examiners' views about the future of driver testing

The interviewers asked questions regarding the future of driver testing. This topic was first addressed with an open question, followed by more specific questions regarding the possibility of automating some parts of the driving test, the use of artificial intelligence to assess candidates, and the use of simulators.

3.3.1. The future of driver testing

The examiners typically mentioned technological developments, such as increased ADAS usage, the increasing presence of electric cars, as well as the replacement of the manual gearbox by an automatic one.

> *"… we should start using more of the assistance systems that are available in cars."* (P4)
>
> *"I expect that … in about ten years, we will have moved to a driving test with automatic transmission."* (P15, Translated)

The main topic of the interview, the data-supported driving test, was often repeated when asked about the future of the driving test.

> *"Indeed, I think the research that you are doing is very positive and that it is indeed moving in that direction that it is becoming data-driven."* (P6, Translated)

Most examiners were not worried about being replaced by a computer.

> *"I am not so afraid about that [being replaced]. [Airplane] pilots are also still needed … I think it is just a shift. You may then need fewer examiners. But something else will take its place. No, I am not afraid of this at all."* (P19, Translated)
>
> *"There are colleagues who are a bit afraid to lose their jobs. But I do not believe it … I think [the systems] will be more supportive systems that can help us establish a verdict."* (P24, Translated)

3.3.2. Automating parts of the driving test

When asked what parts of the driving test could be automated, examiners had difficulty coming up with suggestions. Occasionally, they mentioned parts that they found time-consuming, such as waiting for the candidate to enter a destination in the navigation system. However, no noteworthy suggestions for automation were provided.

3.3.3. Driver assessment by artificial intelligence

The interviewers asked if it would be possible, in the future, to have artificial intelligence assessing candidates partially or completely. Opinions were mixed, but a common ground for all examiners was that they would be difficult to replace.

> *"…. What we actually test of course is whether they have traffic insight, and that word is so elusive, because what is it? But if that were possible [to measure traffic insight], I think it should be possible too [to replace the human examiner]."* (P2, Translated)
>
> *"… you [would] have to collect a huge amount of data, because it is not only the driver and what happens in the car. It is also about the whole environment. Road safety has to do with everything that happens on the road. Everything in his head, in his behavior and in his actions. That is very complex."* (P4)
>
> "*It is always dependent on situations and conditions, and that is the human factor that we add. If someone happens to drive 60 [kph] once, where 50 is allowed, but it is necessary that that speed is driven for a while, … [this does not automatically mean] a failed test.*" (P31, Translated)

While it was agreed upon that it would be complex to achieve complete driver assessment by artificial intelligence, opinions were divided on whether it would be possible or not. Sixteen examiners said it would one day become possible, mentioning a time range between 15 and 40 years.

> *"I think that it is possible, but way in the future. In, like, 40 years or something like that, to do it only with artificial intelligence. Today, no."* (P30)
>
> *"We need a lot of development for that … I think maybe we are 30 years down the road before we can really start to trust the system because it all looks really nice, but there are so many uncertain factors in everything."* (P15, Translated)

On the other hand, eight examiners did not believe it would be feasible, mentioning typically that a computer cannot predict or assess everything.

> *"Making mistakes is human, but the computer systems say 'Yes: correct; this and that went wrong'. Such a computer is, of course, much more black-and-white than we are."* (P37, Translated)
>
> *"Feeling also plays a lot of a role. It just does. And you have those gray cases … and pure data will not be able to make that distinction."* (P11, Translated)

3.3.4. Future role of simulators

Examiners' responses to the question "*Do you think simulators can play a role in the exam*" were mixed. Out of the 37 examiners, 19 saw opportunities for simulators in testing (e.g., "*I surely think so.*"; P26, Translated), and 15 examiners were negative about this ("*Absolutely not.";* P10). Two examiners were ambivalent, and one provided no clear answer.

The examiners saw the benefit of simulators to preselect candidates by testing the basics, to lower the influx of students with poor basic skills. It was noted that simulators cannot fully replace the driving test but that parts of the driving test could be done in the simulator.

> "… *if in a simulator it turns out that somebody really misses all kinds of things, then you do not even have to go to the driving test.*" (P1, Translated)
>
> "*Basic things can already be … [tested]. I think you can capture 80% reasonably well on a simulator.*" (P7, Translated)
>
> "*For example, [you could simulate] a narrow street and there are cars parked on the left and the right side. You can make these standard situations … and see how our candidates react …*" (P4)

A frequently heard argument in favor of introducing simulators in the driving test was that uniform situations can be tested. Currently, outside rush hours, a candidate may pass multiple intersections without any other traffic. Standardized testing may contribute to the fairness of the driving test across districts or times.

> "… *you can more easily present the same situations to people, that allows you to measure more fairly candidate-to-candidate …*" (P2, Translated)
>
> "*You try choosing your route to test all aspects, but sometimes certain situations will not occur. And then I fantasize about simulators, like: 'I would like to have a car coming from the right, now'…*" (P7, Translated)

Examiners noted that limiting factors are simulator sickness and low realism.

> "*When you are already used to driving in a car, and you go to the simulator, you get really nauseous. And you get really sick, and you are not able to drive like you should …*" (P10)
>
> "*You feel nothing, you hear nothing. Yes, you sit still, you do not move; it is very different.*" (P37, Translated)

Negative replies to simulator testing were sometimes followed up by mentioning that there could be a greater role for simulators in driver training. Out of the 18 examiners who were not positive about simulator testing, 10 saw opportunities for training drivers in simulators.

> *"I think it is a very good educational tool. I do not think that it is a useful thing from an assessment point of view."* (P15, Translated)

4. Discussion

This study aimed to assess the views of driving examiners for newly licensed drivers about using data as part of the driving test. The interviews started with questions about the current driving test and the factors that examiners take into consideration when coming to a pass/fail verdict. Subsequently, the interviews went into detail about the why, what, and how of data-driven assessment by discussing examples of presentation and delivery modes of the test results. Simulator-based testing, and offline use of performance data, such as sharing with other examiners, were also addressed in the interviews.

According to the examiners, an important advantage, and a source of job satisfaction, of the current driving test is that examiners have a certain freedom to arrive at a holistic assessment of the candidate's capabilities. For example, the examiner can guide the candidate along an alternative route if the examiner believes that a driving task requires re-assessment. Furthermore, examiners are not obliged to fail an overall competent candidate who made benign errors (CBR, 2020). In the same vein, a stated advantage of the driving test was that the candidates are expected to show independence, for example, by driving to a particular destination themselves with the help of a route navigation device. These characteristics of the driving test correspond to the Goals for Driver Education, which were created some 20 years ago (Hatakka et al., 2002; Keskinen, 2007) and which are increasingly embedded in driver training and testing worldwide (e.g., Alger & Sundström, 2013; Molina et al., 2014; Rodwell et al., 2018; Senserrick et al., 2017). This trend is in line with research showing that safe driving is not attributed to vehicle-handling and maneuvering skills; rather, higher-order skills, such as choosing the appropriate route, insight, and self-reflection, are regarded as essential determinants of safe driving (Gregersen, 1995; Isler et al., 2011; Watson-Brown et al., 2019).

It may be hard for a computer to assess a candidate's higher-order driving skills, for the same reason that automated vehicles have difficulty understanding traffic context and predicting what other road users will do (Rudenko et al., 2020; Vinkhuyzen & Cefkin, 2016). Using Endsley's (1995) terminology: computers may excel at low-level situation awareness (i.e., perception via sensors) but have difficulty achieving high-level situation awareness (comprehension, anticipation of the traffic situation). Consistent with this viewpoint, the examiners pointed out that they cannot rely solely on data for obtaining an overall picture of the candidate. Reliance on data, in a sense, goes against the holistic approach examiners tend to have nowadays. According to the examiners, data should only be used as an aid and should be interpreted in context, for example, by relating the data to surrounding traffic or by combining the data with geographical and real-time traffic information obtained via connected smart-mobility applications (and see Roemer, 2021; and Vissers & Tsapi, 2020, who recommend the integration of smart mobility in the curriculum).

The examiners saw value in measuring proximity to other road users and driving speed (4.4 and 3.9, respectively, on a scale of 1 to 5), which are critical components of safe driving (SWOV, 2012). The measurement of eye movements was considered important as well, while assessments of driver posture or eco-driving were regarded as of less importance. Poor eco-driving is not a reason for failing a candidate and can often be noticed directly from engine sound or dashboard readings without needing supplementary data.

However, the interviews made clear that there are limits to what a driving test can test (be it a data-driven test or not), an observation consistent with the literature. For example, candidates may be susceptible to the looked-but-failed-to-see phenomenon (Herslund & Jørgensen, 2003), make errors because they are nervous (Fairclough et al., 2006), or show rule compliance during the test but reveal themselves as risk-takers once licensed (Baughan et al., 2005).

The interviews showed that examiners are under considerable time pressure and have little time to assess the candidate's driving ability. Of note, as early as 1992, Meijman et al. assessed the workload of Dutch driving examiners and concluded that *"the examiners' job must be characterized as a high stress job"* (pp. 255–256), based on which a recommendation was adopted to reduce the number of driving tests per examiner per day (for a similar study, see Parkes, 1995). Workload and shortage of time were important factors for examiners to accept or reject certain forms of data-driven support. For example, it became clear that, apart from direct warning signals, there is little opportunity to process data *during* the driving test, as examiners are busy monitoring safety, giving instructions, and making sure the candidate is at ease. Also after the driving test, only little time is available, and hence data would need to be available in a straightforward format, or the test structure would have to be changed. At the same time, the examiners emphasized the need to have transparent access to the raw data or graphs, since it is the examiner's role to explain how a verdict is reached. Of note, in France, the test outcome is communicated a few days after the exam (Sécurité Routière, 2021), an approach that would allow the examiners to spend more time analyzing the data before formulating the verdict.

The above findings can be related to the levels of automation proposed by Sheridan (1992, p. 358). On a scale of 1 "*the computer offers no assistance, human must do it all*" to 10 "*the computer decides everything and acts autonomously, ignoring the human*", the examiners would accept Level 2 or 3 at maximum: "*the computer offers a complete set of action alternatives*" and "*narrows the selection down to a few*". In other words, the examiners were favorable towards having access to computer-generated material such as graphs or scores but would not want higher levels of support. Indeed, perhaps the most striking result from the interviews was that the examiners did not want the computer to make the pass/fail decisions for them. Instead, they wanted to use data and video material to clarify and justify their verdict or rule out doubts about the candidate's viewing behavior, headway to the car in front, speed, or braking behavior. This includes the use of data to convince candidates who strongly disagree with the examiner's verdict and who, in some cases, display aggression towards examiners, a problem also noted by others (Foxe, 2020; Roemer, 2021). In a way, the proposed use of data resembles how police patrol uses speed measurements and cameras to show offenders that they violated the traffic rules (Young and Regan, 2007), a concept referred to by some examiners. Body cams have been proposed for the UK driving test to curb violence against examiners (GOV.UK, 2017). Whether body cams are effective in reducing assaults or the number of complaints filed is an ongoing topic of debate (Ariel et al., 2018; Lum et al., 2019). Some examiners expressed concerns that if data were to be shared for evaluation purposes, this could cause an increase in formal complaints.

The scientific literature offers various ideas for introducing higher levels of automation into the driving test. In Fridman et al. (2019), deviant driving was automatically flagged by a computer and then passed to a human supervisor. Such a concept would correspond to Level 5 automation according to Sheridan's (1992) ten-level taxonomy: "*the computer … suggests one [action alternative], and executes that suggestion if the human approves*". The notion of event-triggered data- and camera-based monitoring as part of graduated driver licensing (GDL) has been discussed and studied extensively in the literature (Baker et al., 2020; Klauer et al., 2016; McGehee et al., 2007; Williams & Shults, 2010). Even higher levels of automation are possible, such as in Dubai, where the verdict is supposedly made by a computer with human involvement (e.g., Level 7: "*the computer …executes automatically, then necessarily informs the human*", albeit in a controlled driving range (Government of Dubai, 2019). From the interviews, it became clear that the examiners were hesitant and skeptical about fully automated assessments, noting that computers are unable to make a holistic assessment. The examiners did recognize, however, that automated driver assessments may have a role in specific subtasks, such as special maneuvers or acceleration behavior. In summary, the examiners appeared to be open-minded about the use of data in their current job (Level 2 and 3 automation), while the notion of a fully automated driving test was regarded as unfeasible for the coming decades.

The present interview study concerned the use of in-vehicle technology for assessing driver behavior. A related topic, assessing how drivers use in-vehicle technology, becomes of increasing interest to licensing organizations as well. How drivers of different experience levels change gear was once a topic of considerable academic interest (e.g., Duncan et al., 1991; Shinar et al., 1998), but with the growing popularity of automated gearboxes and electric cars, this component of expertise may disappear, as pointed out by some examiners. Furthermore, newly sold cars contain various ADAS, including blind-spot warning, forward collision warning, adaptive cruise control, lane assist, or other forms of shared control (Oviedo-Trespalacios et al., 2021; Ziebinski et al., 2017). Driving instructors and licensing organizations face growing challenges regarding the training and testing of drivers' interaction with ADAS and automated driving systems (Heikoop et al., 2020; Sturzbecher et al., 2015; Van den Beukel et al., 2021). An increasing body of research aims to examine which training methods are suitable for learning how to interact with assisted and automated driving technology (Ebnali et al., 2019; Manser et al., 2019; Merriman et al., 2021; Noble et al., 2019; Payre et al., 2017; Shaw et al., 2020). For several years in the Netherlands, it is permitted to use ADAS in the driving test (Claesen, 2018), but according to a questionnaire study among driving instructors and examiners, driver assessment of ADAS use is not yet incorporated in driver training and testing in a structured manner (Vlakveld & Wesseling, 2018).

One of the challenges in using ADAS in the driving test is that ADAS availability differs between vehicle models and that different ADAS have different purposes (e.g., comfort/luxury option vs. safety benefits; Tsapi, 2015; Vlakveld & Wesseling, 2018). Similar challenges can be expected in future data-supported driver testing, as variability in vehicles and sensors may compromise the fairness of the assessment. Therefore, attention must be paid to standardization and legislation of data-driven assessment technologies. Regarding legislation, while in the UK, for example, it is not allowed for candidates to record audio or video during the driving test (GOV.UK, 2021), Poland (Kamiński et al., 2008) and Pakistan (Government of Pakistan, 2019) are reported to video-record their driving tests. The examiners argued that the data should be made accessible to the candidates and, provided that privacy is properly taken into consideration, were in favor of sharing data with their employer to improve the quality and uniformity of the driving test. Of note, the Dutch Central Office of Driving Certification already adheres to some open data principles by making the pass rates of all driving schools and examination locations available (CBR, 2021).

A limitation of this study is that it is possible that the examiners' responses were influenced by the familiarity heuristic (Metcalfe et al., 1993). The examiners may have brought up particular possibilities of data-driven assessment because they encountered similar technology in their job (examiners are often seated in modern vehicles and appeared very knowledgeable about ADAS) and may have had difficulty envisioning new ways of data use. In particular, the data type most frequently suggested by the examiners were eye-gaze measurements (Table 2). This may legitimately be a crucial element, as incorrect viewing behavior is a common reason for failing the test (De Winter et al., 2008b). However, many examiners had recently received on-the-job training on eye movements, which may also explain why they brought up this topic. Similarly, the examiners' views about simulators may have been shaped by the fact that simulators are used in driver training in the Netherlands for many years already (De Winter et al., 2008a; Kappé and Van Emmerik, 2005) and are a well-known topic of discussion (United States: Allen et al., 2010; Australia: Rodwell et al., 2019; Norway: Sætren et al., 2018). The examiners regarded simulators as promising for training and screening in standardized conditions but not as a suitable full replacement of the current driving test. Simulator fidelity and simulator sickness in some drivers remain bottlenecks in the acceptance of simulators (De Winter et al., 2012; Kappé & Van Emmerik, 2005).

Another limitation is that the present interviews were conducted with a specific sample: all participants were driving examiners and of Dutch nationality. Future research should assess the views of other participants, such as test candidates (e.g., novice drivers, but also older drivers and professional drivers) as well as examiners and candidates from other nationalities. The Netherlands is a country that adopts a test-led model without obligatory driver training modules and without fixed routes of the driving test. Although the examiners were open-minded about the use of data before and after the driving test, the scope of this research was limited to the driving test itself. Differences in the setup of driver training and testing between countries (e.g., training-led models, multi-phase models) can be expected to lead to different opinions about data use (for overviews of national differences, see Genschow et al., 2014; Helman et al., 2017). Research on other possible data uses, such as whether data could support lifelong learning, should still be performed.

5. Conclusion

Cars are becoming 'computers on wheels', and an increasing number of mobile devices are available that produce driving-related data. These developments raise the question of whether data-driven assessments could have a role in formal driver testing. Interviews were conducted with 37 driving examiners from all testing regions in the Netherlands. The interviews examined if and why examiners would like to use data and what data format would be most useful for them.

It is concluded that examiners are positive about receiving data in the driving test, especially if the data could help them explain their verdict to the candidate. Frequently suggested data types were recordings of the candidates' eye movements and data that describe the car's state in relation to its surroundings, such as speed relative to traffic, distance to surroundings, and position on the road. Examiners were also positive about the use of video fragments, flagged at critical situations. Data should be presented in an easily accessible format, allowing the examiner to obtain an overview in the limited time available between the driving test and the presentation of the verdict. Another key finding was that examiners emphasized the human element in testing drivers and the importance of establishing an overall picture of the candidate.

Our observations are relevant in the context of recently published recommendations stating that the Dutch driving education system needs a fundamental overhaul from a test-led system towards a test- and education-driven system (Roemer, 2021; see Helman et al., 2017 for similar recommendations in a European perspective). For example, it has been recommended that the Netherlands should introduce a modular curriculum and a student monitoring system. The same report recommends conducting experiments with instrumented vehicles to take steps towards a more competency-based assessment (Roemer, 2021). It is expected that the current interview study provides a suitable basis for determining what type of data-driven technology could be used in this experimental phase. Finally, there is a need for knowledge on data-driven assessment in a broader perspective. Future interview studies and experiments could be performed as part of an international consortium that takes into account other target groups, such as truck drivers, as well.

CRediT authorship contribution statement

Tom Driessen: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. **Angèle Picco:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Writing – original draft, Writing – review & editing. **Dimitra Dodou:** Conceptualization, Methodology, Project administration, Supervision, Validation, Writing – review & editing. **Dick de Waard:** Conceptualization, Funding acquisition, Supervision, Writing – review & editing. **Joost de Winter:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Methodology, Project administration, Supervision, Validation, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research is funded by Transitions and Behaviour grant 403.19.243 ("Towards Safe Mobility for All: A Data-Driven Approach"), provided by the Netherlands Organization for Scientific Research (NWO). We would like to thank the Dutch Central Office of Driving Certification (CBR) for their support in recruiting the examiners.

Appendix A. Interview guide

Part 1: Attitudes towards the existing driving exam

1. What are, according to you, the strong aspects of the exam as it is today?
2. What are the flaws in the exam as it is today? If you had the possibility to change anything, what would you modify? How would you do it?
3. Does the driving exam as it is today allow you to assess whether a candidate would drive safely later on?
4. Would you say that the examiner's intuition plays a role in the establishment of a verdict? If yes, to what extent?

Part 2: Attitudes towards the potential data-supported driving exam

(a) **Examiners' opinion about the use of data**
   1. Do you think that the availability of driving data of any form could be of help in the exam?
   2. Do you have any examples of data, even if it sounds strange or impossible to measure?
   3. We will provide you with several examples. For each example, we will ask you to rate this idea on the following scale from 1 to 5, 1 being not useful at all and 5 being very useful. Besides the formal answer on this scale, you are encouraged to share your thoughts, or if you come up with additional ideas.
      Examples:
      a. Video fragments of critical situations
      b. Automatically detected drivers' posture and actions (mirror checking, hands on the wheel)
      c. Distance to detected objects (pedestrians, cyclists, other vehicles)
      d. Deceleration and acceleration (g) scores
      e. Fuel consumption/economic scores
      f. Recordings of speed
      g. Detected failure to follow traffic rules and norms
   4. Can you think of any other data that may be useful?

(b) **Examiners' opinion about the characteristics the data should have**
   1. Now, no matter the type of data collected, when do you think it should be collected? (before, during, after the exam)
   2. When should the data be provided to you? (before, during, after the exam)
   3. How would you envision (/imagine) that the data are presented to you?
   4. With whom do you think the data should be shared? (only you, the candidate, other examiners, data scientists…)
   5. What do you think candidates will think about the use of data to assess them?
   6. Is there a difference to you between driving skills and driving style? Do you think that computers could be good for the two of these?

(c) **Examiners' views about the future of the driving exam**
   1. Is there something that takes a lot of your time during the ride during the driving exam that could be automated?
   2. We talked about how data could help you in your assessment of drivers. Do you think that in the future, an artificial intelligence could assess a driver completely, partially or completely?
      If yes, what would it take for you to rely on this artificial intelligence?
   3. Do you think driving simulators can play a role in the examination?
   4. How do you see the future of driving examinations?

Part 3: General questions

1. How open do you think your organization is to such technological changes? And the examiners themselves?
2. Do you think the pandemic can lead to changes in the setup of the driving exam / the work at CBR?
3. Do you have experience with other training/exams (motorcycle, truck, older drivers, otherspecial domains)? If yes, specify. How do you think that the topic of data-driven assessment applies to that work domain?
4. Do you think the current topic may be useful to driving schools, to implement data in the training?

Appendix B. Code

The Python code for the automated counting of tagged responses and a brief demonstration can be found on GitHub, see https://github.com/tomdries/content-analysis-tools.

Appendix C. Correlation matrix

The means of the seven items listed in Fig. 2 were computed per participant to get an indication of the extent to which participants were positive about the suggested concepts for data-driven assessment. The overall mean of the 37 participants was 3.51 ($SD = 0.67$) on a five-point scale from 1 (*not useful at all*) to 5 (*very useful*).

The mean ratings showed no significant correlations with the participants' age ($r = 0.02$, $p = 0.916$), years of being an examiner ($r = -0.03$, $p = 0.865$), and years of prior experience as a driving instructor ($r = -0.05$, $p = 0.751$). Also, there was no significant difference between the mean rating of males ($M = 3.49$, $SD = 0.70$, $n = 28$) and females ($M = 3.57$, $SD = 0.62$, $n = 9$), $t(35) = -0.29$, $p = 0.771$. Table C1 shows the full correlation matrix among the participants' characteristics and their ratings.

Table C1
Pearson correlation matrix for demographic variables and ratings of the concepts ($n = 37$).

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Age (years)													
2. Gender (0: female, 1: male)	0.15												
3. Years worked as examiner	0.71	0.04											
4. Years worked as instructor	0.13	−0.12	−0.14										
5. Worked as instructor (0: 0 years, 1: > 0 years)	0.10	−0.05	0.06	0.79									
6. No. of driving tests per week	−0.12	0.15	−0.20	0.21	0.16								
7. Distance to objects (1 to 5)	0.09	0.01	0.04	−0.20	−0.23	0.09							
8. Video fragments (1 to 5)	−0.25	0.04	−0.28	−0.03	−0.03	0.05	0.53						
9. Speed (1 to 5)	−0.23	0.08	−0.34	−0.02	−0.04	0.36	0.58	0.54					
10. De-/acceleration (1 to 5)	0.02	−0.07	0.08	−0.13	−0.01	0.19	0.51	0.29	0.52				
11. Rules/norms adherence (1 to 5)	−0.09	−0.03	−0.16	−0.05	−0.14	−0.06	0.47	0.32	0.43	0.22			
12. Posture/actions (1 to 5)	0.26	−0.16	0.15	0.16	0.20	0.01	0.16	0.23	0.22	0.53	0.05		
13. Eco driving (1 to 5)	0.30	−0.10	0.46	0.03	0.06	−0.41	0.24	−0.05	−0.05	0.14	0.05	0.14	
14. Mean rating (1 to 5)	0.02	−0.05	−0.03	−0.05	−0.04	0.08	0.79	0.67	0.75	0.75	0.57	0.56	0.29

References

2todrive, 2021 2todrive. (2021). Begeleid rijden vanaf je 17e [Supervised driving from your 17th]. https://2todrive.nl.

Ahmad, F. H., Zhang, Y., & Qiao, F. (2021). Comparative study of dashcam-based vehicle incident detection techniques. In *Proceedings of the 28th Conference of FRUCT Association. Moscow, Russia* (pp. 539–544). https://doi.org/10.5281/zenodo.4514947

Alger, S., & Sundström, A. (2013). Agreement of driving examiners' assessments–Evaluating the reliability of the Swedish driving test. *Transportation Research Part F: Traffic Psychology and Behaviour, 19*, 22–30. https://doi.org/10.1016/j.trf.2013.02.004

Allen, R. W., Park, G. D., & Cook, M. L. (2010). Simulator fidelity and validity in a transfer-of-training context. *Transportation Research Record: Journal of the Transportation Research Board, 2185*(1), 40–47. https://doi.org/10.3141/2185-06

AmberScript. (2021). Transform your audio and video to text and subtitles. https://www.amberscript.com.

Ariel, B., Sutherland, A., Henstock, D., Young, J., Drover, P., Sykes, J., … Henderson, R. (2018). Paradoxical effects of self-awareness of being observed: Testing the effect of police body-worn cameras on assaults and aggression against officers. *Journal of Experimental Criminology, 14*(1), 19–47. https://doi.org/10.1007/s11292-017-9311-5

Arumugam, S., & Bhargavi, R. (2019). A survey on driving behavior analysis in usage based insurance using big data. *Journal of Big Data, 6*, 86. https://doi.org/10.1186/s40537-019-0249-5

Baker, S. A., Klauer, C., Baynes, P., & Dingus, T. A. (2020). *Parent perceptions of real-time and non-video feedback from the Driver Coach Study (Report No. 20-UM-082)*. Blacksburg, VA: National Surface Transportation Safety Center for Excellence.

Baughan, C., & Simpson, B. (1999). Consistency of driving performance at the time of the L-test, and implications for driver testing. In G. B. Grayson (Ed.), *Behavioural research in road safety IX* (pp. 206–214). Crowthorne: Transport Research Laboratory.

Baughan, C. J., Gregersen, N. P., Hendrix, M., & Keskinen, E. (2005). Towards European standards for testing (Final Report). Brussels: The International Commission for Driving Testing. http://www.cieca.eu/sites/default/files/documents/projects_and_studies/EU_TEST_Project_Final_Report.pdf.

Bergasa, L. M., Almería, D., Almazán, J., Yebes, J. J., & Arroyo, R. (2014). Drivesafe: An app for alerting inattentive drivers and scoring driving behaviors. In *2014 IEEE Intelligent Vehicles Symposium Proceedings. Dearborn, MI* (pp. 240–245). https://doi.org/10.1109/IVS.2014.6856461

Bjørnskau, T. (2003). *Stryk eller stå. En undersøkelse av faktorer som påvirker resultatene av praktisk førerprøve [Pass or fail. A study of factors that affect the results of the driving test] (TØI report No. 662/2003)*. Oslo: Department of Transport Economics.

Boelhouwer, A., Van den Beukel, A. P., Van der Voort, M. C., Verwey, W. B., & Martens, M. H. (2020). Supporting drivers of partially automated cars through an adaptive digital in-car tutor. *Information, 11*, 185. https://doi.org/10.3390/info11040185

Boggs, A. M., Arvin, R., & Khattak, A. J. (2020). Exploring the who, what, when, where, and why of automated vehicle disengagements. *Accident Analysis & Prevention, 136*, 105406. https://doi.org/10.1016/j.aap.2019.105406

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

CBR. (2020). Rijprocedure B [Driving procedure B]. https://www.cbr.nl/nl/voor-rijscholen/nl/partner-in-verkeersveiligheid-1/rijprocedure.htm.

CBR. (2021). Open data. https://www.cbr.nl/nl/over-het-cbr/over/cbr-in-cijfers/open-data.htm.

Chen, B., Zhao, D., & Peng, H. (2017). Evaluation of automated vehicles encountering pedestrians at unsignalized crossings. In *Proceedings of the 2017 IEEE Intelligent Vehicles Symposium. Los Angeles, CA* (pp. 1679–1685). https://doi.org/10.1109/IVS.2017.7995950

Chowdhury, A., Shankaran, R., Kavakli, M., & Haque, M. M. (2018). Sensor applications and physiological features in drivers' drowsiness detection: A review. *IEEE Sensors Journal, 18*(8), 3055–3067. https://doi.org/10.1109/JSEN.2018.2807245

Claesen, R. (2018). Automated driving. The consequences for future driver training and testing. *Zeitschrift für Verkehrssicherheit, 64*, 98–99.

Crinson, L. F., & Grayson, G. B. (2005). Profile of the British learner driver. In G. Underwood (Ed.), *Traffic and transport psychology, theory and application* (pp. 157–170). London: Elsevier.

De Waard, D., & Brookhuis, K. A. (1999). Driver support and automated driving systems: Acceptance and effects on behavior. In M. W. Scerbo, & M. Mouloua (Eds.), *Automation technology and human performance: Current research and trends* (pp. 49–57). Norfolk, VA: Old Dominion University Research Foundation.

De Winter, J. C. F. (2013). Predicting self-reported violations among novice license drivers using pre-license simulator measures. *Accident Analysis & Prevention, 52*, 71–79. https://doi.org/10.1016/j.aap.2012.12.018

De Winter, J. C. F., De Groot, S., Dankelman, J., Wieringa, P. A., Van Paassen, M. M., & Mulder, M. (2008a). Advancing simulation-based driver training: lessons learned and future perspectives. In *Proceedings of the 10th International Conference on Human-Computer Interaction with Mobile Devices and Services. Amsterdam, The Netherlands* (pp. 459–464). https://doi.org/10.1145/1409240.1409314

De Winter, J. C. F., De Groot, S., Van Loenhout, M. J., Van Leeuwen, A., Do, P., Wieringa, P. A., & Mulder, M. (2008b). Feedback on mirror-checking during simulation-based driver training. In J. C. F. De Winter, Advancing simulation-based driver training [Doctoral dissertation] (pp. 153–161). Delft, The Netherlands: Delft University of Technology. http://resolver.tudelft.nl/uuid:d5f04e70-37aa-4a53-8e8d-af17cc402d1e.

De Winter, J. C. F., Dodou, D., Happee, R., & Eisma, Y. B. (2019). Will vehicle data be shared to address the how, where, and who of traffic accidents? *European Journal of Futures Research, 7*(1). https://doi.org/10.1186/s40309-019-0154-3

De Winter, J. C. F., & Kovácsová, N. (2016). How science informs engineering, education, and enforcement: A message for driving instructors. In D. L. Fisher, J. K. Caird, W. J. Horrey, & L. M. Trick (Eds.), *Handbook of teen and novice drivers: Research, practice, policy, and directions* (pp. 31–45). Boca Raton, FL: Taylor & Francis.

De Winter, J. C. F., Van Leeuwen, P. M., & Happee, R. (2012). In *Advantages and disadvantages of driving simulators: a discussion* (pp. 47–50). Noldus Information Technology B.V.: Wageningen.

Duncan, J., Williams, P., & Brown, I. (1991). Components of driving skill: Experience does not mean expertise. *Ergonomics, 34*(7), 919–937. https://doi.org/10.1080/00140139108964835

Ebnali, M., Hulme, K., Ebnali-Heidari, A., & Mazloumi, A. (2019). How does training effect users' attitudes and skills needed for highly automated driving? *Transportation Research Part F: Traffic Psychology and Behaviour, 66*, 184–195. https://doi.org/10.1016/j.trf.2019.09.001

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors, 37*, 32–64. https://doi.org/10.1518/001872095779049543

Fairclough, S. H., Tattersall, A. J., & Houston, K. (2006). Anxiety and performance in the British driving test. *Transportation Research Part F: Traffic Psychology and Behaviour, 9*(1), 43–52. https://doi.org/10.1016/j.trf.2005.08.004

Farag, W. A. (2020). Model-predictive-control complex-path tracking for self-driving cars. *International Journal of Modelling, Identification and Control, 34*, 265–277. https://doi.org/10.1504/IJMIC.2020.111624

Farah, H., Musicant, O., Shimshoni, Y., Toledo, T., Grimberg, E., Omer, H., & Lotan, T. (2014). Can providing feedback on driving behavior and training on parental vigilant care affect male teen drivers and their parents? *Accident Analysis & Prevention, 69*, 62–70. https://doi.org/10.1016/j.aap.2013.11.005

Foxe, K. (2020). More than 50 incidents of aggression by driving test candidates reported. https://www.irishexaminer.com/news/arid-40195446.html.

Fridman, L., Ding, L., Jenik, B., & Reimer, B. (2019). Arguing machines: Human supervision of black box AI systems that make life-critical decisions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Long Beach, CA*. https://doi.org/10.1109/CVPRW.2019.00173

Genschow, J., Sturzbecher, D., & Willmes-Lenz, G. E. (2014). Novice driver preparation—an international comparison (Berichte der Bundesanstalt für Straßenwesen. Mensch und Sicherheit Heft M 234 b). Bergisch Gladbach, Germany: Federal Highway Research Institute.

Giridharan, V. (2019). How Microsoft has quietly started automating driving tests in India using AI. https://www.digit.in/news/car-tech/how-microsoft-has-quietly-started-automating-driving-tests-in-india-using-ai-50854.html.

González, D., Pérez, J., Milanés, V., & Nashashibi, F. (2015). A review of motion planning techniques for automated vehicles. *IEEE Transactions on Intelligent Transportation Systems, 17*, 1135–1145. https://doi.org/10.1109/TITS.2015.2498841

Goodall, N. J. (2021). Comparison of automated vehicle struck-from-behind crash rates with national rates using naturalistic data. *Accident Analysis & Prevention, 154*, 106056. https://doi.org/10.1016/j.aap.2021.106056

Government of Dubai. (2019). RTA announces full automation of drivers testing at the smart yard [Press release]. https://www.rta.ae/wps/portal/rta/ae/home/news-and-media/all-news/NewsDetails/rta+-announces-full-automatio-of-drivers-testing-at-the-smart+yard.

Government of Pakistan. (2019). National guidelines for driver licensing. Islamabad: Ministry of Communications. http://www.roadsafetypakistan.pk/download/national-guidelines-for-driver-licensing.pdf.

GOV.UK. (2017). DVSA aims to stop violence against its staff. https://www.gov.uk/government/news/dvsa-aims-to-stop-violence-against-its-staff.

GOV.UK. (2021). Filming or recording driving tests: what you're allowed to do. https://www.gov.uk/government/publications/filming-or-recording-driving-tests/filming-or-recording-driving-tests.

Graesser, A. C., Chipman, P., Haynes, B. C., & Olney, A. (2005). AutoTutor: An intelligent tutoring system with mixed-initiative dialogue. *IEEE Transactions on Education, 48*(4), 612–618. https://doi.org/10.1109/TE.2005.856149

Gregersen, N. P. (1995). What should be taught? Basic vehicle control skills or higher order skills? In H. S. Simpson (Ed.), *New to the road: reducing the risks for young motorists. Proceedings of the First Annual International Conference of the Youth Enhancement Service. Los Angeles, CA* (pp. 103–114).

Handel, P., Ohlsson, J., Ohlsson, M., Skog, I., & Nygren, E. (2013). Smartphone-based measurement systems for road vehicle traffic monitoring and usage-based insurance. *IEEE Systems Journal, 8*(4), 1238–1248. https://doi.org/10.1109/JSYST.2013.2292721

Hatakka, M., Keskinen, E., Gregersen, N. P., Glad, A., & Hernetkoski, K. (2002). From control of the vehicle to personal self-control; broadening the perspectives to driver education. *Transportation Research Part F: Traffic Psychology and Behaviour, 5*(3), 201–215. https://doi.org/10.1016/S1369-8478(02)00018-9

Heikoop, D. D., Calvert, S. C., Mecacci, G., & Hagenzieker, M. P. (2020). A practitioner's view of driver training for automated driving from driving examiners: A focus group discussion. In *2020 Forum on Integrated and Sustainable Transportation Systems. Delft, The Netherlands* (pp. 14–19). https://doi.org/10.1109/FISTS46898.2020.9264869

Helman, S., Vlakveld, W., Fildes, B., Oxley, J., Fernández-Medina, K., & Weekley, J. (2017). Study on driver training, testing and medical fitness (Final report). Retrieved from https://ec.europa.eu/transport/road_safety/sites/default/files/dl_study_on_training_testing_med_fitness.pdf.

Herslund, M.-B., & Jørgensen, N. O. (2003). Looked-but-failed-to-see-errors in traffic. *Accident Analysis & Prevention, 35*(6), 885–891. https://doi.org/10.1016/S0001-4575(02)00095-7

Hickman, J. S., & Hanowski, R. J. (2011). Use of a video monitoring approach to reduce at-risk driving behaviors in commercial vehicle operations. *Transportation Research Part F: Traffic Psychology and Behaviour, 14*(3), 189–198. https://doi.org/10.1016/j.trf.2010.11.010

Isler, R. B., Starkey, N. J., & Sheppard, P. (2011). Effects of higher-order driving skill training on young, inexperienced drivers' on-road driving performance. *Accident Analysis & Prevention, 43*(5), 1818–1827. https://doi.org/10.1016/j.aap.2011.04.017

Kamiński, T., Wojciechowski, A., Nowacki, G., Mitraszewska, I., Kamińska, E., & Filipek, P. (2008). Video register picture evaluation to support drivers training and examination. *Journal of KONES Powertrain and Transport, 15*, 225–232. https://www.infona.pl//resource/bwmeta1.element.baztech-article-BUJ5-0034-0027.

Kappé, B., & Van Emmerik, M. L. (2005). Mogelijkheden van rijsimulatoren in de rijopleiding en het rijexamen [The possibilities of driving simulators for initial driver training and testing] (Report No. TNO-DV3 2005 C114). Soesterberg: TNO Defense and Security.

Karvonen, H., Kujala, T., & Saariluoma, P. (2006). In-car ubiquitous computing: driver tutoring messages presented on a head-up display. In *Proceedings of the 2006 IEEE Intelligent Transportation Systems Conference. Toronto, Ontario* (pp. 560–565). https://doi.org/10.1109/ITSC.2006.1706800

Kashevnik, A., Lashkov, I., & Gurtov, A. (2019). Methodology and mobile application for driver behavior analysis and accident prevention. *IEEE Transactions on Intelligent Transportation Systems, 21*(6), 2427–2436. https://doi.org/10.1109/TITS.2019.2918328

Keskinen, E. (2007). What is GDE all about and what it is not. In W. Henriksson, T. Stenlund, A. Sundstrom, & M. Wiberg (Eds.), *Proceedings from The GDE-Model as a Guide in driver training and testing* (pp. 3–13). Umea, Sweden: Umea University.

Klauer, S. G., Ehsani, J., & Simons-Morton, B. (2016). Using naturalistic driving methods to study novice drivers. In S. G. Klauer, J. Ehsani, & B. Simons-Morton (Eds.), *Handbook of teen and novice drivers* (pp. 429–440). CRC Press.

Lechner, G., Fellmann, M., Festl, A., Kaiser, C., Kalayci, T. E., Spitzer, M., & Stocker, A. (2019). A lightweight framework for multi-device integration and multi-sensor fusion to explore driver distraction. In P. Giorgini, & B. Weber (Eds.), *Advanced Information Systems Engineering. CAiSE 2019* (pp. 80–95). Cham: Springer. https://doi.org/10.1007/978-3-030-21290-2_6.

Lima, P. F., Pereira, G. C., Mårtensson, J., & Wahlberg, B. (2018). Experimental validation of model predictive control stability for autonomous driving. *Control Engineering Practice, 81*, 244–255. https://doi.org/10.1016/j.conengprac.2018.09.021

Litman, T. (2021). *Autonomous vehicle implementation predictions. Implications for transport planning.* Victoria, British Columbia: Victoria Transport Policy Institute.

Lum, C., Stoltz, M., Koper, C. S., & Scherer, J. A. (2019). Research on body-worn cameras: What we know, what we need to know. *Criminology & Public Policy, 18*(1), 93–118. https://doi.org/10.1111/1745-9133.12412

Magaña, V. C., & Muñoz-Organero, M. (2015). Artemisa: A personal driving assistant for fuel saving. *IEEE Transactions on Mobile Computing, 15*, 2437–2451. https://doi.org/10.1109/TMC.2015.2504976

Manser, M. P., Noble, A. M., Machiani, S. G., Shortz, A., Klauer, S. G., Higgins, L., & Ahmadi, A. (2019). Driver training research and guidelines for automated vehicle technology (Final Report No. 01-004). Blacksburg, VA: Virginia Tech Transportation Institute.

Marin-Plaza, P., Hussein, A., Martin, D., & De la Escalera, A. (2018). Global and local path planning study in a ROS-based research platform for autonomous vehicles. *Journal of Advanced Transportation, 2018*, 1–10. https://doi.org/10.1155/2018/6392697

Marti, E., De Miguel, M. A., Garcia, F., & Perez, J. (2019). A review of sensor technologies for perception in automated driving. *IEEE Intelligent Transportation Systems Magazine, 11*(4), 94–108. https://doi.org/10.1109/MITS.2019.2907630

Mase, J. M., Agrawal, U., Pekaslan, D., Mesgarpour, M., Chapman, P., Torres, M. T., & Figueredo, G. P. (2020). Capturing uncertainty in heavy goods vehicles driving behaviour. In *Proceedings of the 2020 IEEE 23rd International Conference on Intelligent Transportation Systems. Rhodes, Greece*. https://doi.org/10.1109/ITSC45102.2020.9294378

McGehee, D. V., Raby, M., Carney, C., Lee, J. D., & Reyes, M. L. (2007). Extending parental mentoring using an event-triggered video intervention in rural teen drivers. *Journal of Safety Research, 38*(2), 215–227. https://doi.org/10.1016/j.jsr.2007.02.009

Meijman, T. F., Mulder, G., Van Dormolen, M., & Cremer, R. (1992). Workload of driving examiners: A psychophysiological field study. In H. Kragt (Ed.), *Enhancing industrial performance* (pp. 245–258). London, UK: Taylor & Francis.

Merriman, S. E., Plant, K. L., Revell, K. M., & Stanton, N. A. (2021). Challenges for automated vehicle driver training: A thematic analysis from manual and automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour, 76*, 238–268. https://doi.org/10.1016/j.trf.2020.10.011

Metcalfe, J., Schwartz, B. L., & Joaquim, S. G. (1993). The cue-familiarity heuristic in metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 851–861. https://doi.org/10.1037/0278-7393.19.4.851

Michelaraki, E., Katrakazas, C., Yannis, G., Filtness, A., Talbot, R., Hancox, G., … Taveira, R. (2021). Post-trip safety interventions: State-of-the-art, challenges, and practical implications. *Journal of Safety Research, 77*, 67–85. https://doi.org/10.1016/j.jsr.2021.02.005

Michon, J. A. (Ed.). (1993). *Generic intelligent driver support*. London, UK: CRC Press. https://doi.org/10.1201/9781003208952.

Microsoft. (2021). HAMS: Harnessing AutoMobiles for Safety. https://www.microsoft.com/en-us/research/project/hams.

Molina, J. G., García-Ros, R., & Keskinen, E. (2014). Implementation of the driver training curriculum in Spain: An analysis based on the Goals for Driver Education (GDE) framework. *Transportation Research Part F: Traffic Psychology and Behaviour, 26*, 28–37. https://doi.org/10.1016/j.trf.2014.06.005

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to cousineau (2005). *Tutorials in Quantitative Methods for Psychology, 4*, 61–64. https://doi.org/10.20982/tqmp.04.2.p061

Musicant, O., & Lotan, T. (2016). Can novice drivers be motivated to use a smartphone based app that monitors their behavior? *Transportation Research Part F: Traffic Psychology and Behaviour, 42*, 544–557. https://doi.org/10.1016/j.trf.2015.10.023

Mynttinen, S., Koivukoski, M., Hakuli, K., & Keskinen, E. (2011). Finnish novice drivers' competences–Successful driving test candidates 2000–2009 evaluated by driving examiners. *Transportation Research Part F: Traffic Psychology and Behaviour, 14*(1), 66–75. https://doi.org/10.1016/j.trf.2010.10.001

Nambi, A. U., Bannur, S., Mehta, I., Kalra, H., Virmani, A., Padmanabhan, V. N., Bhandari, R., & Raman, B. (2018). HAMS: Driver and driving monitoring using a smartphone. In *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking. New Delhi, India* (pp. 840–842). https://doi.org/10.1145/3241539.3267723

Noble, A. M., Klauer, S. G., Doerzaph, Z. R., & Manser, M. P. (2019). Driver training for automated vehicle technology–knowledge, behaviors, and perceived familiarity. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 63*(1), 2110–2114. https://doi.org/10.1177/1071181319631249

Olweus, D. (1958). *Undersökning rörande bilinspektörernas bedömning av de praktiska körkortsproven [Investigation concerning the driving examiners' assessments of practical driving license tests]*. Stockholm, Sweden: Institute of Psychotechnics, Stockholm University.

Oviedo-Trespalacios, O., Tichon, J., & Briant, O. (2021). Is a flick-through enough? A content analysis of Advanced Driver Assistance Systems (ADAS) user manuals. *PLOS ONE, 16*. https://doi.org/10.1371/journal.pone.0252688

Parkes, K. R. (1995). The effects of objective workload on cognitive performance in a field setting: A two-period cross-over trial. *Applied Cognitive Psychology, 9*(7), S153–S171. https://doi.org/10.1002/acp.2350090710

Payre, W., Cestac, J., Dang, N. T., Vienne, F., & Delhomme, P. (2017). Impact of training and in-vehicle 'task performance on manual control recovery in an automated car. *Transportation Research Part F: Traffic Psychology and Behaviour, 46*, 216–227. https://doi.org/10.1016/j.trf.2017.02.001

Pugnetti, C., & Elmer, S. (2020). Self-assessment of driving style and the willingness to share personal information. *Journal of Risk and Financial Management, 13*, 53. https://doi.org/10.3390/jrfm13030053

Ramzan, M., Khan, H. U., Awan, S. M., Ismail, A., Ilyas, M., & Mahmood, A. (2019). A survey on state-of-the-art drowsiness detection techniques. *IEEE Access, 7*, 61904–61919. https://doi.org/10.1109/ACCESS.2019.2914373

Ranft, B., & Stiller, C. (2016). The role of machine vision for intelligent vehicles. *IEEE Transactions on Intelligent Vehicles, 1*(1), 8–19. https://doi.org/10.1109/TIV.2016.2551553

Rangesh, A., & Trivedi, M. M. (2019). No blind spots: Full-surround multi-object tracking for autonomous vehicles using cameras and lidars. *IEEE Transactions on Intelligent Vehicles, 4*(4), 588–599. https://doi.org/10.1109/TIV.2019.2938110

Rodwell, D., Hawkins, A., Haworth, N., Larue, G. S., Bates, L., & Filtness, A. (2018). A mixed-methods study of driver education informed by the Goals for Driver Education: Do young drivers and educators agree on what was taught? *Safety Science, 108*, 140–148. https://doi.org/10.1016/j.ssci.2018.04.017

Rodwell, D., Hawkins, A., Haworth, N., Larue, G. S., Bates, L., & Filtness, A. (2019). What do driver educators and young drivers think about driving simulators? A qualitative draw-and-talk study. *Transportation Research Part F: Traffic Psychology and Behaviour, 62*, 282–293. https://doi.org/10.1016/j.trf.2019.01.008

Roemer, E. (2021). Van rijles naar rijonderwijs. Advies verbetering autorijscholenbranche [From driving lessons to driving education. Advice for improving the driving School industry]. The Hague, The Netherlands: Ministry of Infrastructure and Water Management. https://www.rijksoverheid.nl/documenten/rapporten/2021/04/14/bijlage-1-van-rijles-naar-rijonderwijs-advies-verbeteren-autorijscholenbranche.

Ropelato, S., Zünd, F., Magnenat, S., Menozzi, M., & Sumner, R. (2018). Adaptive tutoring on a virtual reality driving simulator. *International SERIES on Information Systems and Management in Creative EMedia, 2017*, 12–17. https://doi.org/10.3929/ethz-b-000195951

Rudenko, A., Palmieri, L., Herman, M., Kitani, K. M., Gavrila, D. M., & Arras, K. O. (2020). Human motion trajectory prediction: A survey. *The International Journal of Robotics Research, 39*(8), 895–935. https://doi.org/10.1177/0278364920917446

Sætren, G. B., Pedersen, P. A., Robertsen, R., Haukeberg, P., Lindheim, C., & Rasmussen, M. (2018). Simulator training in driver education – potential gains and challenges. In S. Haugen, A. Barros, C. van Gulijk, T. Kongsvik, & J. E. Vinnem (Eds.), *Safety and reliability – Safe societies in a changing world* (pp. 2045–2049). London, UK: Taylor & Francis Group.

Schoettle, B. (2017). Sensor fusion: A comparison of sensing capabilities of human drivers and highly automated vehicles (Report No. SWT-2017-12). Ann Arbor, MI: University of Michigan.

Schöner, H.-P., Pretto, P., Sodnik, J., Kaluza, B., Komavec, M., Varesanovic, D., … Antona-Makoshi, J. (2021). A safety score for the assessment of driving style. *Traffic Injury Prevention, 22*(5), 384–389. https://doi.org/10.1080/15389588.2021.1904508

Sécurité Routière. (2021). Résultats du permis de conduire [Driving license results]. https://www.securite-routiere.gouv.fr/resultats-du-permis-de-conduire#/step-connexions.

Segers, J. (2014). *Analysis techniques for racecar data acquisition.* Warrendale, PA: SAE International.

Senserrick, T., McRae, D., Wallace P, W., De Rome, L., Rees, P., & Williamson, A. (2017). Enhancing higher-order skills education and assessment in a graduated motorcycle licensing system. *Safety, 3*(2), 14. https://doi.org/10.3390/safety3020014

Shanly, C., Ieti, M., Warren, I., & Sun, J. (2018). BackPocketDriver – A mobile app to enhance safe driving for youth. In Proceedings of the 30th International Conference on Software Engineering and Knowledge Engineering (pp. 246–245). Redwood City, CA. https://doi.org/10.18293/SEKE2018-011.

Shaw, E., Large, D., & Burnett, G. (2020). *Driver training for future automated vehicles: Introducing CHAT (CHeck, Assess, Takeover)*. Nottingham, UK: University of Nottingham.

Sheridan, T. B. (1992). *Telerobotics, automation, and human supervisory control*. Cambridge, MA: MIT Press.

Shimshoni, Y., Farah, H., Lotan, T., Grimberg, E., Dritter, O., Musicant, O., … Omer, H. (2015). Effects of parental vigilant care and feedback on novice driver risk. *Journal of Adolescence, 38*, 69–80. https://doi.org/10.1016/j.adolescence:2014.11.002

Shinar, D., Meir, M., & Ben-Shoham, I. (1998). How automatic is manual gear shifting? *Human Factors, 40*(4), 647–654. https://doi.org/10.1518/001872098779649346

Shladover, S. E. (2016). The truth about "self-driving" cars. *Scientific American, 314*, 52–57. https://doi.org/10.1038/scientificamerican0616-52

Sikander, G., & Anwar, S. (2018). Driver fatigue detection systems: A review. *IEEE Transactions on Intelligent Transportation Systems, 20*(6), 2339–2352. https://doi.org/10.1109/TITS.2018.2868499

Sturzbecher, D., Mörl, S., & Kaltenbaek, J. (2015). *Optimisation of the practical driving test (Berichte der Bundesanstalt für Straßenwesen. Mensch und Sicherheit Heft M 243 b)*. Potsdam, Germany: University of Potsdam.

SWOV. (2012). SWOV Fact sheet: Headway times and road safety. https://www.swov.nl/sites/default/files/publicaties/gearchiveerde-factsheet/uk/fs_headway_archived.pdf.

SWOV. (2016). SWOV Fact sheet: 18- to 24-year-olds: young drivers. https://www.swov.nl/en/facts-figures/factsheet/18-24-year-olds-young-drivers.

Tabone, W., De Winter, J., Ackermann, C., Bärgman, J., Baumann, M., Deb, S., … Stanton, N. A. (2021). Vulnerable road users and the coming wave of automated vehicles: Expert perspectives. *Transportation Research Interdisciplinary Perspectives, 9*, 100293. https://doi.org/10.1016/j.trip.2020.100293

Terán, J., Navarro, L., Quintero, M. C. G., & Pardo, M. (2020). Intelligent driving assistant based on road accident risk map analysis and vehicle telemetry. *Sensors, 20*, 1763. https://doi.org/10.3390/s20061763

Tsapi, A. (2015). Introducing Advanced Driver Assistance Systems (ADAS) into drivers' training and testing: The young learner drivers' perspective [MSc thesis Delft University of Technology]. http://resolver.tudelft.nl/uuid:1c8f1bb7-c68e-4596-b341-a4f3bb70cdd9.

Tummala, G. K., Das, T., Sinha, P., & Ramnath, R. (2019). SmartDashCam: automatic live calibration for DashCams. In *Proceedings of the 18th International Conference on Information Processing in Sensor Networks. Montreal, Quebec* (pp. 157–168). https://doi.org/10.1145/3302506.3310397

Van den Beukel, A. P., Van Driel, C. J., Boelhouwer, A., Veders, N., & Heffelaar, T. (2021). Assessment of driving proficiency when drivers utilize assistance systems—the case of Adaptive Cruise Control. *Safety, 7*, 33. https://doi.org/10.3390/safety7020033

Vavouranakis, P., Panagiotakis, S., Mastorakis, G., & Mavromoustakis, C. X. (2017). Smartphone-based telematics for usage based insurance. In C. Mavromoustakis, G. Mastorakis, & C. Dobre (Eds.), *Advances in mobile cloud computing and big data in the 5G era. Studies in big data* (pp. 309–339). Cham: Springer. https://doi.org/10.1007/978-3-319-45145-9_13.

Vinkhuyzen, E., & Cefkin, M. (2016). Developing socially acceptable autonomous vehicles. *Ethnographic Praxis in Industry Conference Proceedings, 2016*(1), 522–534. https://doi.org/10.1111/1559-8918.2016.01108

Vissers, J., & Tsapi, A. (2020). Kwaliteitsimpuls rijscholen. Onderzoek naar een rijscholenregister en kwaliteitscontrolesysteem [Quality impulse driving schools. Research into a driving school register and quality control system] (Report No. BG9103-RHD-ZZ-XX-RP-Z-0001). The Hague, The Netherlands: Ministry of Infrastructure and Water Management.

Vlakveld, W. P., & Wesseling, S. (2018). ADAS in het rijexamen: Vragenlijstonderzoek onder rijschoolhouders en rijexaminatoren naar moderne rijtaakondersteunende systemen in de rijopleiding en het rijexamen voor rijbewijs B [ADAS in the driving test: Questionnaire study among driving school owners and driving examiners into modern driving task support systems in driver education and the driving test for driving license B] (Report No. R-2018-20). The Hague, The Netherlands: SWOV – Institute for Road Safety Research.

Wassink, I., Van Dijk, B., Zwiers, J., Nijholt, A., Kuipers, J., & Brugman, A. (2006). In the Truman Show: Generating dynamic scenarios in a driving simulator. *IEEE Intelligent Systems, 21*, 28–32. https://doi.org/10.1109/MIS.2006.97

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. *Safety Science, 136*, 105172. https://doi.org/10.1016/j.ssci.2021.105172

Watson-Brown, N., Scott-Parker, B., & Senserrick, T. (2019). Association between higher-order driving instruction and risky driving behaviours: Exploring the mediating effects of a self-regulated safety orientation. *Accident Analysis & Prevention, 131*, 275–283. https://doi.org/10.1016/j.aap.2019.07.005

Williams, A. F., & Shults, R. A. (2010). Graduated driver licensing research, 2007–present: A review and commentary. *Journal of Safety Research, 41*(2), 77–84. https://doi.org/10.1016/j.jsr.2010.03.002

Young, K. L., & Regan, M. A. (2007). Intelligent Transport Systems to support Police enforcement of road safety laws (ATSB Research and Analysis Report No. 2007-02). Canberra City, Australia: Australian Transport Safety Bureau.

Zahabi, M., Park, J., Razak, A. M. A., & McDonald, A. D. (2020). Adaptive driving simulation-based training: Framework, status, and needs. *Theoretical Issues in Ergonomics Science, 21*(5), 537–561. https://doi.org/10.1080/1463922X.2019.1698673

Ziebinski, A., Cupek, R., Grzechca, D., & Chruszczyk, L. (2017). Review of advanced driver assistance systems (ADAS). *AIP Conference Proceedings, 1906*, 120002. https://doi.org/10.1063/1.5012394