



Article

Towards the Validation of an Observational Tool to Detect Impaired Drivers—An Online Video Study

Tanja Baertsch ^{1,*} , Marino Menozzi ¹  and Signe Maria Ghelfi ²

¹ Department of Health Sciences and Technology, ETH Zürich, 8092 Zurich, Switzerland; mmenozzi@ethz.ch

² Zurich State Police, Airport Division—Research and Development, Zurich-Airport, 8058 Kloten, Switzerland; signe@gmx.ch

* Correspondence: tanja.baertsch@hest.ethz.ch

Abstract: Abuse of alcohol and other drugs is a major risk factor at work. To reduce this risk, workplace drug testing is performed in transportation and other industries. VERIFY, an observational method, is one of the key elements in a procedure adopted by the police of the canton of Zurich, Switzerland, for detecting impaired drivers. The observational method has been successfully applied by adequately trained police officers since 2014. The aim of this study is to examine the interrater reliability of the observational method, the effect of training in use of the method, and the role of having experience in the police force and traffic police force on the outcome when rating a driver's impairment. For this purpose, driver impairment in staged road traffic controls presented in videos was rated by laypeople ($n = 81$), and police officers without ($n = 146$) and with training ($n = 172$) in the VERIFY procedure. In general, the results recorded for police officers with training revealed a moderate to very good interrater reliability of the observational method. Among the three groups, impaired drivers were best identified by officers with training (ranging between 82.6% and 89.5% correct identification). Trained officers reported a higher impairment severity of the impaired drivers than the other two groups, indicating that training increases sensitivity to signs of impairment. Our findings also suggest that online video technology could be helpful in identifying impaired drivers. Trained police officers could be connected to a road traffic control to make observations via live video. By this method efficiency and reliability in detecting abuse of alcohol and other drugs could be improved. Our findings also apply to workplace drug testing in general.

Keywords: accident prevention; workplace drug testing; impaired driving; video-based training



Citation: Baertsch, T.; Menozzi, M.; Ghelfi, S.M. Towards the Validation of an Observational Tool to Detect Impaired Drivers—An Online Video Study. *Int. J. Environ. Res. Public Health* **2022**, *19*, 7548. <https://doi.org/10.3390/ijerph19127548>

Academic Editor: Paul B. Tchounwou

Received: 19 April 2022

Accepted: 17 June 2022

Published: 20 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Across cultures, alcohol, other psychoactive substances, medicinal drugs, sleepiness, and their combined effects are serious issues that impact occupational safety [1–8] and lead to economic costs, expenditures on health care, and costs related to law enforcements [9,10]. In particular, substance abuse and sleepiness may lead to accidents, adversely affecting people and environments, such as in transportation (aviation, train, and road traffic in general) and nuclear power plants. To keep work safe, workplace drug testing (WDT) has been introduced in various industries [11,12], for example, the aviation sector [13], to identify drug abuse at an individual level. Similarly, objective roadside measurements, such as breathalyzers for estimating blood alcohol content or oral drug screening devices, are used in police traffic controls to foster traffic safety. Because of their high specificity to particular drugs, roadside drug tests are blind to other drugs, such as new psychoactive substances, most medicinal drugs, and combinations of substances [14–16].

In addition to objective measurements, the so-called impairment tests are applied as a subjective means to detect impairment by evaluating the behavior and appearance of the examinee. With subjective methods, effects on examinees' behavior and appearance are recorded and evaluated rather than measuring the cause itself, which is a major advantage

of subjective methods over objective approaches. Thus, the issues with screening tests measuring only specific substances can be overcome. However, currently widely used subjective methods are time-consuming. In police traffic control, for instance, the drug recognition expert evaluation used in the United States, which includes an interview, recording clinical signs, and performing psychomotor tests, takes 90 min [17] and the Standardized Field Sobriety Tests, including the gaze nystagmus test, walk and turn, and one-leg stand, takes approximately 10 min [18]. In addition to being time-consuming, impairment tests are biased by subjectivity and depend on expertise in correctly administering them.

Overall, subjective methods can be helpful in complementing WDT and in assessing fitness for work. Considering its successful application since 2014, the subjective observational method called 'VERIFY' (a German acronym: *Verfahren zur Identifikation von Fahruntfähigkeit*—procedure for identification of driving impairment) is a promising tool to overcome various problems of the currently widely used strategies of impairment detection. VERIFY was developed by police officers, who are experts in the detection of impaired drivers, with the collaboration of scientists in pharmaceutical science. VERIFY has proven to be highly effective as the results of observational evaluations were in good agreement with the results of blood and urine analyses and/or the opinion of experts of the Institute of Forensic Medicine on the overall case. According to Boll 2019 [19], in 85% of VERIFY cases where impairment was detected, the driving impairment was confirmed by analytical results for the blood and urine samples of the examinees or/and the expert opinions on the overall cases.

VERIFY consists of a checklist in which observations of driver's appearance and behavioral factors related to ocular functions, cognition, motor functions, and others are systematically reported. The decision of whether a driver is impaired is based on the totality of the evidence and not simply on one sign of impairment. Before applying VERIFY in practice, police officers must complete a training in which they apply VERIFY during road traffic controls under the guidance of experts. A detailed description of the VERIFY method is given in Appendix A.1, and a sample of the checklist is shown in Figure 1.

The present study aims to investigate the interrater reliability of the checklist (Figure 1) and the agreement of the rating with the state of impairment of a driver. Furthermore, the effect of the training is examined by comparing ratings of driving impairment between police officers who underwent the training and police officers without training. The latter group obtains reasonable suspicion by objective methods and by observation of signs of impairment but without having the possibility to conduct and report observations as systematically as in the VERIFY procedure. A third group consisting of laypeople was included in the comparison as a control. It is hypothesized that trained police officers are more reliable in identifying impaired drivers than police officers without training. Furthermore, as it is widely assumed that experts perceive risks differently from laypeople and that their judgments are more truthful than those made by laypeople [20], it is hypothesized that laypeople perceive driving impairment as less severe than police officers. Finally, the effect of experience in the police force and in the traffic police force on rating outcomes is investigated.

Alcohol odor <input type="checkbox"/> yes <input type="checkbox"/> no	Appearance <input type="checkbox"/> normal/inconspicuous <input type="checkbox"/> unkempt <input type="checkbox"/> neglected	Physical signs <input type="checkbox"/> none <input type="checkbox"/> sweating <input type="checkbox"/> trembling <input type="checkbox"/> restlessness <input type="checkbox"/> vomiting	Reaction <input type="checkbox"/> normal/inconspicuous <input type="checkbox"/> delayed <input type="checkbox"/> extremely delayed
Cannabis odor <input type="checkbox"/> yes <input type="checkbox"/> no	Command of the German language <input type="checkbox"/> yes <input type="checkbox"/> no <input type="checkbox"/> conditionally	Speech <input type="checkbox"/> normal/inconspicuous <input type="checkbox"/> stuttering <input type="checkbox"/> slurred <input type="checkbox"/> mumbling	Response <input type="checkbox"/> normal/inconspicuous <input type="checkbox"/> sleepy <input type="checkbox"/> arousable <input type="checkbox"/> sleeping/unconscious
Orientation <input type="checkbox"/> normal/inconspicuous <input type="checkbox"/> disoriented	Gait <input type="checkbox"/> normal/inconspicuous <input type="checkbox"/> insecure <input type="checkbox"/> sluggish <input type="checkbox"/> unsteady <input type="checkbox"/> staggering	Mood/Behaviour <input type="checkbox"/> normal/inconspicuous <input type="checkbox"/> excited <input type="checkbox"/> strangely cheerful <input type="checkbox"/> impassive	<input type="checkbox"/> lack of emotional detachment <input type="checkbox"/> provocative <input type="checkbox"/> aggressive <input type="checkbox"/> whining
Getting out of the car <input type="checkbox"/> normal/inconspicuous <input type="checkbox"/> unbalanced <input type="checkbox"/> has to hold on the vehicle			

Pupil template: values in mm

12	11	10	9	8	7	6	5	4	3	2	1

Eyes <input type="checkbox"/> normal/inconspicuous <input type="checkbox"/> conjunctiva reddened <input type="checkbox"/> watery/glassy <input type="checkbox"/> restless	Pupils right ca. mm _____ left ca. mm _____ <input type="checkbox"/> prompt reaction to light <input type="checkbox"/> sluggish reaction to light	Lighting condition at control location <input type="checkbox"/> daylight <input type="checkbox"/> twilight <input type="checkbox"/> night/street lighting <input type="checkbox"/> night/room lighting
--	---	---

Found on the person or in the vehicle (narcotics, narcotics paraphernalia, alcohol, drugs, etc.)

Other observations (abnormalities when driver approaching, on the person or in the vehicle)

Behaviour during official act (not only physical examination)

consistent increasingly conspicuous increasingly inconspicuous

Duration of the official act from _____ (time)
to _____ (time)

Conclusion of the police officer (reasons for urgent suspicion of driving impairment)

Figure 1. The VERIFY checklist consists of several items, which can be ticked if police officer thinks they apply to driver, and several open-ended questions. The VERIFY checklist is originally in German.

2. Materials and Methods

Our study required participants to use the observational tool, VERIFY, for rating the impairment of drivers in a police traffic control. To avoid variations caused by the direct observation of real-road traffic controls, ratings were performed based on recorded videos showing staged controls of impaired and unimpaired drivers. Video observations have been reported to be a reliable substitute for direct observations in a variety of applications, such as observing posture [21], assessing mental health [22] or depression [23], or detecting lies when conducting police interrogations [24].

Due to ethical considerations and to overcome limitations in recruitment, the roles of drivers in the videos showing staged controls were played by unimpaired actors. According to a study by Paquette and Prince [25], drug recognition experts who were trained solely through simulated evaluations with unintoxicated actors are as effective in detecting intoxicated persons in the field as police officers trained conventionally with intoxicated

persons. The strategy of using simulated patients (SP) has successfully been used in medical education since the 1950s [25] and was found to have many advantages over using real patients, such as SP are available at any time, the use of SP avoids mistreatment of real patients, and SP are evaluated without embarrassment of the student [26]. Conclusively, there is evidence that simulated impairment detections are an effective alternative to impairment detections with real impaired persons. Video observations do not require observers to be present at a specific place and at a specific time. Instead, they enable them to efficiently collect observational data using a standalone web-based method.

2.1. Participants

Three groups were involved in the study: laypeople, police officers without VERIFY education, and VERIFY-trained police officers (i.e., police officers who underwent VERIFY education).

Laypeople were recruited via a participant panel provided by the Department of Psychology at the University of Zurich. Laypeople were selected according to demographical, cultural, and professional criteria. They had to be 18 years or older, possess a driving license, understand Swiss German, and never have worked in the police force before. Police officers with and without VERIFY education were recruited by the traffic police of the canton of Zurich following the same criteria regarding age, driving license, and language as used in laypeople. In total, 151 police officers without VERIFY education, 173 VERIFY-trained police officers, and 82 laypeople were recruited.

To improve the quality of the collected data, a seriousness check was performed retrospectively [27]. As the total duration of all videos was 15 min, participants who took less than 15 min to complete the study were excluded because they did not watch all videos to the end. In total, seven participants were excluded from the analyses. For the three groups, the sociodemographics of the included participants and the duration for completing the ratings are presented in Table 1.

Table 1. Sociodemographics and duration in minutes and seconds for completing the rating by groups.

		Group Laypeople	Without VERIFY Education	With VERIFY Education
Age	Min–Max	19–79	25–62	24–59
	M ^a (SD) ^b	31.01 (13.14)	40.76 (11.22)	37.27 (7.71)
Gender	Female	54 (66.7%)	42 (28.8%)	31 (18%)
	Male	27 (33.3%)	104 (71.2%)	141 (82%)
	Total	81 (100%)	146 (100%)	172 (100%)
Duration (min)	Min	16:37	15:17	18:7
	Max	49:58	58:6	88:15
	M ^a (SD) ^b	21:85 (5:47)	26:58 (7:54)	32:23 (10:77)

^a Mean ^b Standard deviation.

The self-reported experience levels of police officers in traffic police or in other police forces are presented in Table 2.

Table 2. Self-reported years of experience of police officers in traffic police or in other police forces for police officers with and without education in the VERIFY procedure.

		Group without VERIFY Education	with VERIFY Education
Experience in police force [years]	M ^a (SD) ^b	15.06 (11.82)	11.83 (7.56)
	N [total]	146	172
Experience in traffic police force [years]	M ^a (SD) ^b	4.44 (5.24)	6.08 (5.27)
Currently works for traffic police	N	29	81
Worked for traffic police in the past	N	30	37
	N [total]	59	118

^a Mean ^b Standard deviation.

2.2. Material

Six videos of staged road traffic controls, each lasting approximately two and a half minutes, were recorded during a VERIFY training session in which police officers practiced the VERIFY procedure. The videos demonstrated two police officers stopping a driver, conversing, and examining their pupils. To simulate pupils of impaired drivers, medical educational videos displaying the effects of drugs on pupils were edited into the videos. The size of the edited pupils was the same for the left and right eyes in all videos. The role of all actors in the videos were played by real police officers, and the officer playing the impaired driver pretended to be impaired. There were three scenes with unimpaired drivers and three scenes with impaired drivers. A short description of the content of each video is presented in Table 3.

Table 3. Details of videos showing staged traffic controls.

	Impairment of Stopped Driver	Signs of Impairment
Video 1	Under influence of alcohol or another depressant	Slow reaction; unfocused; confusion; imbalance; unsteady gait
Video 2	Feeling very ill, might be under influence of medicinal drug	Medicinal drug box in car; slow reaction; sleepy; unsteady gait
Video 3	Under influence of a stimulant	Hyper behavior; large pupils; inappropriately cheerful; lack of emotional detachment
Video 4	No impairment	
Video 5	No impairment	
Video 6	No impairment	

2.3. Procedure

The online study was programmed using Unipark survey software. A link to the online experiment was sent to the participants via e-mail. The experiment could be conducted on a computer at any place or time.

Prior to the experiment, participants were provided with an introductory text explaining the aim of the study, which is to investigate how the assessment of a driver's ability to drive differs between police officers and laypeople. The text also explained that the task was to watch six videos of staged road traffic controls, rate the driver's impairment severity, and state the reasons for the suspicion if driving impairment was suspected.

Participants were asked to conduct the experiment alone and without interruption. Before the experiment started, participants answered sociodemographic questions (gender, age). In addition, laypeople were asked for their last educational achievement whereas police officers were asked for years of experience as a police officer and as a road traffic police officer. The data was collected anonymously. A numerical identifier was associated

to the data collected with the questionnaire. After that, participants watched a short trial video to check the video and audio quality of their computer. Subsequently, the six videos were presented in a randomized order. Participants were instructed to watch all six videos in a row without repetition or interruption. From the recorded data, it was checked whether the videos were watched only once and without interruption.

After each video, participants in all groups were first asked whether they thought the driver in the video was unimpaired (dependent variable I). This question was answered with a yes or no. Subsequently, participants were asked to indicate the extent to which they agreed or disagreed with the following statements:

- (I) The driver complied with the instructions of the police.
- (II) The driver showed conspicuous behavior.
- (III) The driver showed abnormalities in movement.

A five-point Likert scale ranging from “strongly disagree” to “strongly agree” was used to report the extent of agreement with the statements (II)–(IV).

In cases where the driver was classified as impaired (variable I), participants belonging to the group without VERIFY education or laypeople were further asked to state at least one reason for their suspicion. The group with VERIFY-trained police officers reported their observations by completing the VERIFY checklist (Figure 1). The checklist consists of 49 categorical items (e.g., slurred speech, slow reaction, the smell of alcohol, etc.), two interval-scaled items (right and left pupil sizes), and three open-ended questions. The checklist includes a pupil template with different pupil sizes in mm to help estimate pupil size. Participants were informed that the items about the odor and the lightning condition at the control location should not be considered, as these could not be evaluated in the staged videos.

The study protocol was approved by the ethical committee of ETH Zürich (approval No. 2021-N-03).

2.4. Data Analysis

The data were analyzed using SPSS version 26.0 (IBM Corp., Armonk, NY, USA, 2019). Pearson’s χ^2 was calculated to examine the variation of the assessment of impairment (dependent variable 1) across groups. For significance ($p \leq 0.05$), the Bonferroni post hoc test was conducted. Cramer’s V was calculated as an effect size measure [28].

On the evaluation results of each video, Fleiss’ kappa [29] was run to investigate the interrater reliability of the VERIFY checklist. The guidelines of Landis and Koch (1977) [30] were used to classify the level of agreement with respect to the value of Fleiss’ kappa. This interrater reliability calculation excluded the open-ended questions and the items about odor and lighting conditions at the control location. As pupil size may vary depending on the types of drugs and the types of drugs influencing the staged impaired driver were not specified in our study, ‘left pupil size’ and ‘right pupil size’ were not used for interrater evaluations. Boxplot distributions of sizes were used as a means for evaluating the accuracy in reporting pupil size.

Cronbach’s alpha was used to determine the internal consistency coefficient of the three statements on driver impairment (see dependent variables II–IV in Section 2.3: Procedure). A score of driving impairment severity was formed based on the three dependent variables II–IV. For this purpose, the coding of the dependent variable II was first reversed, and then the means of agreement with the statements (variables II–IV) were calculated for each video.

To investigate whether the three groups of participants differed in estimating the severity of driving impairment, one-way analyses of variances (ANOVA) with Bonferroni post hoc tests were conducted using the score of impairment recorded for each video. As the homogeneity of variance was only met in results related to videos 1, 2, 3, and 6, Welch ANOVAs were applied to the results of videos 4 and 5.

The evaluation of the reported state of the driver (impaired/not impaired) and the severity of driving impairment included an examination of whether having experience in the police force and in the traffic police force affected the assessment of driving impairment.

For this purpose, ANOVA was conducted with the groups of police officers with experience in road traffic policing and police officers without experience in road traffic policing.

3. Results

The results are presented in the following sequence: interrater reliability, reported pupil sizes, detection of the impairment of the driver, the role of work experience in traffic police on reported state of driver (impaired/not impaired), assessment of the impairment severity, and the role of work experience in traffic police on reported severity of the impairment.

3.1. Interrater Reliability of the VERIFY Checklist

Police officers who thought a driver was impaired completed the VERIFY checklist. The number of completed ratings for each video is listed in Table 4. Fleiss' kappa revealed moderate to very good agreement between officers' assessments (Table 4).

Table 4. Interrater reliability of the checklist based on responses of officers who underwent education in VERIFY. Details on rated impairments are listed in Table 3.

Video	Impairment	n	Fleiss' Kappa	p-Value	95% CI
1	Impaired	146	0.48 ^c	<0.01	[0.48, 0.48]
2	Impaired	142	0.48 ^c	<0.01	[0.48, 0.48]
3	Impaired	154	0.73 ^b	<0.01	[0.73, 0.73]
4	Unimpaired	9	0.90 ^a	<0.01	[0.90, 0.91]
5	Unimpaired	18	0.75 ^b	<0.01	[0.75, 0.75]
6	Unimpaired	16	0.67 ^b	<0.01	[0.60, 0.60]

CI = confidence interval, n = number of raters, ^a very good reliability, ^b good reliability, ^c moderate reliability according to Landis & Koch (1977).

3.2. Reported Pupil Sizes

Figure 2 displays the estimated sizes of left and right pupils for each video, which were reported by officers who had undergone education in VERIFY. In seven cases, police officers rated the left and right pupil sizes differently.

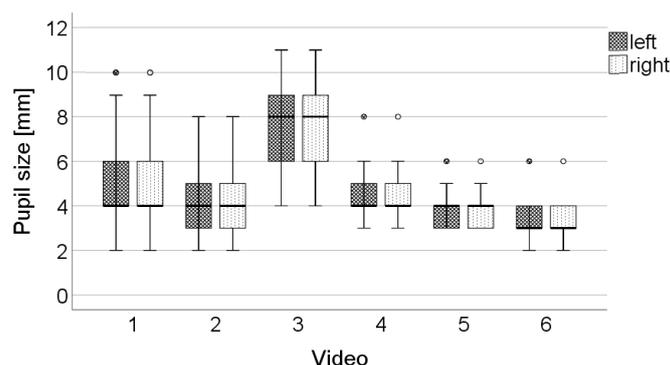


Figure 2. Boxplot representation of left and right pupil sizes per video. n: number of VERIFY-trained police officers who estimated pupil size. Videos 1–3 display impaired driver, and in videos 4–6, the driver is not impaired.

3.3. Detection of Impairment of the Driver

Table 5 reports the results for the detection of impairment for each rated video separately. For each rated video, the results of the χ^2 test are listed in the last column.

Table 5. Frequency analyses and χ^2 tests for rating in driving impairment.

Video Driver's Condition	Group	Correct Assessment	Incorrect Assessment	Total	χ^2 Tests of Independence
Video 1 <i>impaired</i>	Laypeople	45 (55.6%)	36 (44.4%)	81 (100%)	$\chi^2(2) = 29.26$ $p < 0.01$ $\phi = 0.27^a$
	withoutVerify	119 (81.5%)	27 (18.5%)	146 (100%)	
	withVerify	146 (84.9%)	26 (15.1%)	172 (100%)	
Video 2 <i>impaired</i>	Laypeople	52 (64.2%)	29 (35.8%)	81 (100%)	$\chi^2(2) = 12.75$ $p < 0.01$ $\phi = 0.18^a$
	withoutVerify	120 (82.2%)	26 (17.8%)	146 (100%)	
	withVerify	142 (82.6%)	30 (17.4%)	172 (100%)	
Video 3 <i>impaired</i>	Laypeople	67 (82.7%)	14 (17.3%)	81 (100%)	$\chi^2(2) = 5.10$ $p = 0.08$ $\phi = 0.11^a$
	withoutVerify	118 (80.8%)	28 (19.2%)	146 (100%)	
	withVerify	154 (89.5%)	18 (10.5%)	172 (100%)	
Video 4 <i>unimpaired</i>	Laypeople	79 (97.5%)	2 (2.5%)	81 (100%)	$\chi^2(2) = 6.03$ $p = 0.06$ $\phi = 0.12^a$
	withoutVerify	131 (89.7%)	15 (10.3%)	146 (100%)	
	withVerify	163 (94.8%)	9 (5.2%)	172 (100%)	
Video 5 <i>unimpaired</i>	Laypeople	72 (88.9%)	9 (11.1%)	81 (100%)	$\chi^2(2) = 2.78$ $p = 0.25$ $\phi = 0.08^a$
	withoutVerify	122 (83.6%)	24 (16.4%)	146 (100%)	
	withVerify	154 (89.5%)	18 (10.5%)	172 (100%)	
Video 6 <i>unimpaired</i>	Laypeople	72 (88.9%)	9 (11.1%)	81 (100%)	$\chi^2(2) = 2.71$ $p = 0.37$ $\phi = 0.07^a$
	withoutVerify	125 (85.6%)	21 (14.4%)	146 (100%)	
	withVerify	156 (90.7%)	16 (9.3%)	172 (100%)	

^a Small effect size [28]. ϕ = effect size (Cramer's V).

For the ratings on videos 1 and 2, χ^2 revealed a significant difference between the three groups of participants. For the ratings collected on Video 1, the post hoc test displayed that the expected and observed frequencies significantly differed in the group of laypeople as well as in the group with VERIFY-trained police officers.

Laypeople were more likely to let an impaired driver continue driving than expected, whereas VERIFY-trained police officers were more likely to detect an impaired driver than expected. For the results collected in Video 2, the post hoc χ^2 test revealed that the expected and observed frequencies significantly differed in the laypeople group. Laypeople were more likely to let the impaired driver continue driving than expected.

Role of Work Experience in Traffic Police on the Reported State of Driver (Impaired/ Not Impaired)

When comparing the assessment of whether a driver was impaired or not between police officers with work experience and ratings of police officers without work experience in traffic police analogous, as done in the previous Section 3.3, no significant differences in ratings were found.

3.4. Assessment of Impairment Severity

Cronbach's α analysis including the three components of the impairment severity scale (see Section 2.3) revealed a good reliability of the components ($\alpha = 0.84$). Differences in the assessment of impairment severity between the groups are reported in Table 6.

Table 6. Comparison of reported impairment severity between groups.

Video	Group	N	Impairment Score [Mean across All Participants and Variables II–IV (SD)]	One-Way ANOVA/Welch Test
Video 1 <i>impaired</i>	Laypeople	81	2.95 (0.98)	F(2, 396) = 41.98 $p < 0.01$ $n_p^2 = 0.18^a$
	withoutVerify	146	3.72 (0.88)	
	withVerify	172	4.06 (0.88)	
	Total	399	3.71 (0.99)	
Video 2 <i>impaired</i>	Laypeople	81	2.77 (0.73)	F(2, 396) = 21.31 $p < 0.01$ $n_p^2 = 0.10^b$
	withoutVerify	146	3.35 (0.75)	
	withVerify	172	3.42 (0.79)	
	Total	399	3.26 (0.80)	
Video 3 <i>impaired</i>	Laypeople	81	2.97 (0.80)	F(2, 396) = 20.52 $p < 0.01$ $n_p^2 = 0.10^b$
	withoutVerify	146	3.06 (0.83)	
	withVerify	172	3.54 (0.77)	
	Total	399	3.25 (0.84)	
Video 4 <i>unimpaired</i>	Laypeople	81	1.16 (0.37)	Welch's F(2, 201.54) = 1.11 $p = 0.33$ $n_p^2 = 0.005^c$
	withoutVerify	146	1.17 (0.54)	
	withVerify	172	1.11 (0.32)	
	Total	399	1.14 (0.42)	
Video 5 <i>unimpaired</i>	Laypeople	81	1.52 (0.69)	Welch's F(2, 198.42) = 1.06 $p = 0.35$ $n_p^2 = 0.005^c$
	withoutVerify	146	1.57 (0.68)	
	withVerify	172	1.48 (0.55)	
	Total	399	1.53 (0.63)	
Video 6 <i>unimpaired</i>	Laypeople	81	1.46 (0.58)	F(2, 396) = 0.20 $p = 0.82$ $n_p^2 = 0.001^c$
	withoutVerify	146	1.41 (0.62)	
	withVerify	172	1.42 (0.67)	
	Total	399	1.42 (0.63)	

^a Large effect [28]. ^b Medium effect [28]. ^c Small effect [28]. n_p^2 = effect size (eta squared).

The Bonferroni post hoc test of impairment severity rating in Video 1 revealed that the observed impairment severity was significantly different between all groups ($p < 0.01$). The estimated impairment severity increased from laypeople ($M = 2.95$, $SD = 0.98$) to the group without VERIFY education ($M = 3.72$, $SD = 0.88$) to the group with VERIFY-trained police officers ($M = 4.06$, $SD = 0.88$). The Bonferroni post hoc test of Video 2 revealed that laypeople ($M = 2.77$, $SD = 0.73$) differed significantly in their rating from the rating in the VERIFY-trained group ($M = 3.42$, $SD = 0.79$) and the group without training ($M = 3.35$, $SD = 0.75$, $p < 0.01$). The post hoc analysis of impairment severity rating in Video 3 revealed a significant difference ($p < 0.01$) between the group of trained police officers ($M = 3.54$, $SD = 0.77$) and laypeople ($M = 2.97$, $SD = 0.80$). There was also a significant difference in ratings between groups with and without training ($M = 3.06$, $SD = 0.83$).

Role of Work Experience in Traffic Police on the Reported Severity of Impairment

When comparing impairment severity rated by police officers with work experience in traffic police to ratings of police officers without work experience in traffic police analogous, as done in the previous Section 3.4, no significant differences in ratings were found.

4. Discussion

Our study aimed to validate an observational tool (VERIFY) for the detection of impairment of drivers presented on staged videos of a police traffic control. In particular, we evaluated the interrater reliability of the tool and how the perceived impairment level as assessed with the tool varied across laypeople, police officers without VERIFY education, and VERIFY-trained police officers. Additionally, we investigated whether the outcome

when rating a driver's impairment depended on whether police officers had experience in the traffic police force.

Our findings revealed a moderate to very good interrater reliability of the observational tool. High Fleiss' kappa scores suggest that different evaluators observed behavior and appearance of the examinee similarly, indicating certain objectivity of the tool despite the subjective approach. However, a large variation in reported pupil sizes was observed (Figure 2) indicating that the estimation of pupil size using the size template alone is problematic. This finding is consistent with results reported by Olson et al. [31], who analyzed interrater reliability of subjective assessment of pupil size and reactivity. According to their results, the interrater reliability of both, size and reactivity estimation was low. Olson et al. [31] suggested considering automated pupillometry for a reliable assessment of pupil size and reactivity. An additional cause for the large variation in reported pupil sizes in the present study is that pupil sizes were reported from memory because the VERIFY checklist was filled in after the observation was completed. It is also likely that it is more difficult to estimate pupil size in a video setting with a short duration than in a real police traffic control, which takes more time than the duration of the video.

The results of the study indicate that identifying unimpaired drivers is easy, as there were no significant differences between the groups in assessing an unimpaired driver.

In one of the three videos showing an impaired driver, the rating difference between the groups was not significant ($p = 0.08$). The video showed a person under the influence of a stimulant. The person was very excited and had extremely large pupils. All groups predominantly suspected impairment, potentially because, in the general population, the effects of stimulant drugs that make a person hyperactive and cause large pupils are widely known. In the other two videos of impaired drivers, significant differences in impairment detection were observed. From these findings, we can postulate that laypeople are more likely to let a person who feels very ill and who might be under the influence of a medicinal drug continue driving (Video 2). This underestimation of diseases and the influence of medicinal drugs on driving behavior are consistent with previous studies [3,4]. Unfortunately, the use of objective screening devices to compensate for a lack of experience in laypeople is not feasible, as actual roadside screening devices cannot detect impairments caused by diseases or medicinal drugs.

In Video 1, a person who was sleepy and confused and who reacted very slowly was assessed as impaired more often by trained police officers than by police officers without training and by laypeople. In other words, trained police officers perceived the trait more clearly than the other two groups. The beneficial effect of training in reporting observed behavior was demonstrated in the study by Angkaw et al. [32], in which videotaped role-play interactions of participants with alcohol use disorders were rated. According to their results, observational training improves the interrater reliability and reduces the discrepancy in rating scores between the raters. Laypeople might not classify sleepy behavior as dangerous and therefore are more likely to let an impaired driver continue driving when clear indications of impairment are absent. The Swiss Council for Accident Prevention (BFU) presumes that the dark figure of accidents caused by sleepiness is as high as 10–20% [33], indicating that sleepy driving is strongly underestimated by the population. The effectiveness of the observational tool in identifying sleepiness is a major advantage over commonly used impairment detection methods.

While other widely used impairment tests, such as the drug recognition expert evaluation test, would have taken up to 90 min to form the suspicion of driving impairment, VERIFY enables detection of a driver's impairment in a shorter time. From the present experience of the police of the canton of Zurich, VERIFY-trained police officers using the tool take approximately five minutes to form an opinion on a driver's condition.

Our results also reveal a variation in the perception of the level of impairment across the three groups. In all videos of impaired drivers, laypeople perceived the impairment as less severe than VERIFY-trained police officers, and the differences between the two groups were significant. This finding is in line with the results reported by Rodriguez-

Garzon et al. [34], who found that volunteer firefighters revealed a lower level of risk perception than professionals. On the other hand, trained police officers rated the level of impairment in unimpaired drivers lower than in laypeople, although the differences were not significant. Taken together, the differences in rating the level of impairment in impaired and unimpaired drivers suggest that trained police officers perform a more accurate rating compared to laypeople, which is also supported by the findings of Hahm et al. [20], who found that expert judgments are more accurate than others.

In contrast to the effect of training in applying the observational tool, officers with work experience in the traffic police force did not detect impairment significantly differently from police officers without experience in the traffic police force. This outcome suggests that skills in the use of the tool are acquired immediately or in a short time after officers receive the training.

Limitations

This study has several limitations. For example, gender distribution differed between the groups. Most of the laypeople were female, whereas in the groups of police officers, the greater proportion was male. This should be kept in mind when interpreting the comparisons between the groups since a driver's appearance or behavior could be perceived differently depending on gender. Hall et al. (2016) [35] reviewed research on gender differences in interpersonal accuracy and concluded that females excel at remembering others' appearances and non-verbal behaviors, and they have more extensive knowledge of the meaning and use of non-verbal communication. This leads to the assumption that females tend to be better at assessing driving ability. However, training can improve accuracy of judgment of others' emotions, personality traits, status, and intentions [36]. As police officers are trained to make such judgments and they need these skills in their daily work, it is assumed that the gender difference in the police groups is negligible. If the gender distribution in the laypeople group had been similar to that in the police groups, laypeople might have assessed driving ability less accurately than in the present study due to the smaller proportion of females. This suggests that if the gender distribution has been similar in the groups, the difference in the assessment of driving ability between the police officers and the laypeople would have been greater than in the present study.

Furthermore, 40% of the group without training and 69% with training had traffic police experience. The remaining proportion worked for the police but not as traffic police officers. One aim of the analysis was to test whether VERIFY training leads to a better result in assessing driving impairment than the group without VERIFY training. One could argue that the comparison is biased because the group without training included people who did not have experience in traffic policing. However, to reduce this bias, the χ^2 and ANOVA tests were conducted with laypeople, police officers with experience in road traffic policing, and police officers without experience in road traffic policing. The results revealed no significant effect of experience on the expected frequency of rating the impairment correctly or assessing impairment severity. Furthermore, observation skills are required not only in the traffic police but also in many other areas of the police service. Therefore, it is very likely that the police officers of the group without training and with no experience as traffic police officers also need these skills in their daily work. It can be assumed that their skills are comparable to those of traffic police officers without any specific training in assessing driving impairment. Thus, this distributional inequality can be neglected.

One of the main limitations is that the traffic controls were presented in videos. Thus, police officers could not conduct the traffic controls themselves or interact with the drivers. Evaluating traffic control on the computer without interacting with the driver is not the same as conducting a real face-to-face control. The observer might recognize more evidence of driving impairment on the computer, as the attention only needs to be focused on the driver's behavior and appearance and not on conducting the conversation. It could also be that raters would have assessed impairments better if they could conduct the control by themselves and interact directly with the driver. Based on the results of Hartwig et al. [24]

and Ihlebæk et al. [37], it can be assumed that there would have been either no difference or a deterioration in the impairment assessment if the participants had conducted the road traffic control by themselves or observed it live.

Moreover, it could be argued that it would have to be proven that any staged behavior of impaired and unimpaired drivers represents what they really mean to define. The actors in the videos imitated the symptoms that often occurred from experience. Therefore, drivers' behaviors in the videos are close to reality, and it can be assumed that the behaviors represent what they mean to define.

Furthermore, the interrater agreement and correct answer rate might depend on the obviousness of the acted behaviors. The results related to the videos of impaired drivers indicate that the interrater agreement might depend on the obviousness of the acted behaviors. The hyperactive person with large pupils shown in Video 3 is clearly impaired, and Cronbach's α is higher than in the other two videos in which the signs of impairment are less apparent. However, the interrater reliabilities in the less apparent cases were still moderate. In terms of the correct answer rate, generally, the less apparent signs of impairment, the more difficult it is to identify an impaired driver. This is evident in the group of laypeople, whose correct answer rate is much higher in Video 3 than in the other two videos. However, there were no clear differences in the correct assessment between the three videos showing impaired drivers in the groups of police officers. This indicates that the correct answer rate does not necessarily depend on the obviousness of the acted behaviors.

Another limitation is that the videos did not allow participants to see all the details, such as the interior of the car, which could indicate a possible inability to drive (e.g., open alcohol bottles). However, the main goal of this study was to investigate whether the behavior and appearance were reported similarly by trained police officers, and this purpose was fulfilled in this study.

5. Conclusions

This study investigated the interrater reliability of an observational tool named VERIFY, which is used for detecting impaired drivers during road traffic controls. We assessed differences in rating driving impairment and its severity by laypeople, police officers trained to use the observational tool, and police officers without training. The results of our study revealed moderate to very good interrater reliability. The assessment of impaired drivers was significantly more reliable in police officers than in laypeople. Contrary to our expectations, trained police officers did not identify impaired drivers significantly better than police officers without training. However, trained police officers perceived the traits more clearly than those without training. This is particularly noticeable in cases where drivers do not display apparent impairment. In real-road traffic controls, drivers rarely display apparent impairments as in the videos used in this study. Thus, we assume that training brings an advantage over using the procedure without any training, especially in more difficult cases. With respect to the training in using the presented observational tool, our study indicates that training is successful, as trained police officers perceive driving impairment as more severe than police officers without training.

A combined application of objective screening methods for drug testing with the observational tool could help improve the reliability of drug testing through triangulation of the results recorded with the two methods. In addition, the tool could help reveal impairment by drugs for which no on-site objective drug testing devices are available.

Similarly as in traffic controls, the observational tool could also be applied to workplace drug testing (WDT) in general for coping with issues of an increasing demand for WDT. As listed by the US Substance Abuse and Mental Health Services Administration (SAMHSA), WDTs are applied in various occasions, such as pre-employment tests, for-cause and reasonable suspicion tests, and post-accident and post-treatment tests. Furthermore, the number of people taking drugs has been on the rise [38,39]. Therefore, there is a need for an accurate, effective, and quick method for identifying impaired workers. The observational tool presented in this study can help improve efficiency and effectiveness in workplace drug

detection to cope with the increasing number of WDTs. The major point of criticism of the observational tool, namely subjectivity, could be mitigated in this study by demonstrating its good interrater reliability.

Our study supports the use of videos to familiarize police officers with the observational procedure. In the future, observational impairment evaluations could make use of video technology to remotely include experts trained in using the observational tool, thereby increasing the efficiency of testing.

Lastly, the findings in this study indicate that the population needs to be made more aware of the dangers of working under the influence of medicinal drugs or sleepiness.

In conclusion, the presented observational tool will improve efficiency and effectiveness in road traffic controls as well as in workplace drug testing and will therefore contribute to a safer work environment.

Author Contributions: Conceptualization, T.B.; methodology, T.B., M.M. and S.M.G.; validation, T.B.; formal analysis, T.B.; investigation, T.B.; resources, T.B., M.M. and S.M.G.; data curation, T.B.; writing—original draft preparation, T.B.; writing—review and editing, T.B., M.M. and S.M.G.; visualization, T.B.; supervision, M.M.; project administration, T.B.; funding acquisition, M.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Zurich Cantonal Police, Switzerland.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Ethics Committee of ETH Zürich (protocol code 2021-N-03, 11 February 2021).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Appendix A.1 VERIFY—Method for Identifying Impaired Drivers

The observation method called ‘VERIFY’ has been in use in the canton of Zurich, Switzerland, since 2014. This is an observation technique in which a specially educated police officer records the appearance and behavioral functions of the driver, such as motor skills, attention, or pupil size, using a protocol. The observations are not made during typical sobriety tests, but they are made during the routine protocol and last approximately five minutes. Therefore, only a little additional time is required to identify a driving impairment, benefiting both the police and the drivers. Police work is not significantly hindered, and unimpaired drivers are not detained for too long.

The VERIFY protocol in which the police officers record drivers’ appearances and behavioral functions includes a checklist consisting of several categories (e.g., speech, gait, and response), each including several items (e.g., speech: stuttering, slurred, and mumbling) and several open-ended questions. Based on the totality of the evidence recorded in the protocol, the officer decides whether there are sufficient indications of impaired driving. VERIFY focuses on indications of impairment of basic functions and reserve capacity. The latter is assessed by giving the driver several tasks at once (e.g., opening the window and turning off the lights). A significant advantage of the VERIFY procedure over other methods is that VERIFY records driving impairment regardless of its cause, and therefore, police officers are not required to decide whether a driver is under the influence of psychoactive substances, sleepiness, diseases, or a combination of these.

Suppose there is a clear suspicion of impaired driving after completing the VERIFY process, a prosecutor decides whether the observed indicators are strong enough to take the driver to the Zurich Institute of Forensic Medicine (ZIFM) at the University of Zurich or to a hospital for taking blood and urine samples and for a medical examination. When

in certain cases, results of urine and blood samples and the examinations do not confirm the initial suspicion of driving impairment, police officers receive feedback on the results of the analyses.

Several approaches have been pursued to control for interrater variabilities in performing the VERIFY procedure. First, a VERIFY checklist specifying clear observation traits is provided to police officers. Furthermore, officers undergo training in which they are instructed how to apply the method during a two-day course. The course ends with a real-road traffic control in which the police officers have to apply what they have learned while a VERIFY trainer assists them. Additionally, police officers regularly perform refresher courses. As a further measure to control for interrater variability, police officers who repeatedly fail in their initial suspicion must attend follow-up training.

Previous data analysis of VERIFY cases recorded in 2020 revealed that the VERIFY process is sensitive to detecting driving impairment. In 92% of the cases where VERIFY-trained police officers detected signs of impairment, the forensic toxicological analysis at the ZIFM at the University of Zurich was either positive for a medicinal or illicit drug or the observed symptoms were confirmed by a medical doctor to be associated with an impairing condition. However, in 28% of those cases, the analytical result, together with the observations of police officers, was insufficient to confirm the presence of impaired driving at the time of the control with the certainty required in court. This does not mean that police officers wrongfully suspected drivers of impaired driving in 28% of cases. A disagreement between the police and the forensic expert's opinion can easily be explained, for example, the fact that a substance was found below a legal threshold because too much time had passed between the control and the sampling and back-calculation of concentration to the time of the event is not possible for medicinal and illicit drugs. Before VERIFY, the Zurich police used a widespread oral fluid drug test that detected only some common drugs, such as cannabis, cocaine, and opiates. In cases where the oral fluid drug test detected at least one substance, the ZIFM was able to confirm the corresponding substance in a sufficient driving-relevant concentration in only about 46% of the blood samples [19]. Consequently, 54% of drivers who tested positive for a drug during traffic control were falsely accused. Thus, since the introduction of VERIFY, the proportion of drivers for whom the expert from the ZIFM found the evidence insufficient to punish them for impaired driving has been reduced from 54% to 28%. This retrospective analysis supports the validity of VERIFY.

Appendix A.2 The Theory behind VERIFY

Police training is the process of imparting the knowledge and skills required for integration into the profession. The contents of training reflect professional qualities, skills, and knowledge [40]. The nature and demands of policing change over time, and thus police officers' education must be adapted to new challenges in the field. A major challenge is, for example, that about 50 to over a hundred new psychoactive substances are developed annually [41]. This leads to the first of three innovative aspects of VERIFY. VERIFY-trained police officers do not use roadside drug tests because they have the disadvantage of not being sensitive to all types of impairments. Police officers learn to identify driving impairments based on their general signs (e.g., disorientation, speech, and gait) independent of the cause of impairment. This presupposes that there are general signs that indicate driving impairment. Before the general signs of impairment can be discussed, driving impairment must be defined. There exists a vague legal definition of when a driver is not allowed to drive a vehicle. The Swiss Road Traffic Act defines any driver as impaired who does not have the required physical or mental capacity due to the influence of alcohol, narcotics, medication, or other reasons [42]. From the point of view of traffic medicine, there are three important functions needed to maintain the required physical or mental capacity for safe driving: vision, cognition, and motor function [43]. If these functions are affected, the risk of accidents increases [44]. Consequently, signs indicating any impairment in vision, cognition, and motor function can be defined as general signs of driving impairment. Regarding the VERIFY checklist, almost all categories of items are related to one of those

functions (Table A1). Although the VERIFY checklist is based on the experience of experts in drug recognition, the content aligns with the theory.

Table A1. Categorization of VERIFY checklist items.

Vision	Cognition	Motor Function	Others
Eyes Pupil size Pupil reaction to light	Orientation Response Mood/Behavior	Getting out of the car Gait Physical signs Reaction Speech	Alcohol odor Cannabis odor Appearance Command of the German language

The second innovative aspect of VERIFY is that police officers focus not only on signs of impairment of basic functions but are also trained to test reserve capacity [45]. Reserve capacity is required when the demand for a situation increases and the additional required performance can no longer be processed within the framework of the normal performance capacity. If both the normal performance capacity and the reserve capacity are exhausted, the performance limit is reached. When this happens, drivers are not able to react to difficult and unpredictable situations, such as when an animal suddenly crosses the street [42]. The literature has additionally demonstrated the importance of reserve capacity by revealing that drivers spend approximately half of their driving time engaged in secondary tasks unrelated to driving, requiring a high level of cognitive effort. As impaired cognitive function caused by, for example, psychoactive substances negatively affects cognitive capacity, drivers are less likely to perform multiple tasks simultaneously, also reducing their ability to deal with difficult and unpredictable situations [46,47]. VERIFY-trained police officers assess reserve capacity by giving the driver several tasks at once (e.g., opening the window and turning off the lights) and observing whether all tasks are accomplished.

A further goal of the conversation between the police officer and driver is to put the driver in a situation in which automatic behavior patterns take over. At the beginning of the control, a driver is likely to be nervous and stressed, as there is a potential risk of being legally punished. In stressful situations, the sympathetic nervous system is highly activated, and the body's fight or flight response is stimulated. The sympathetic nervous system prepares for an appropriate reaction in dangerous, stressful, or challenging situations, whereas the parasympathetic nervous system is activated after a stressful situation and produces a relaxed feeling [48]. Due to this physiological process, it is suspected that a driver with a highly activated sympathetic nervous system can control bodily actions and thus hide an impairment to a certain extent. Conclusively, the third innovative aspect of VERIFY is that police officers intend to use small talk to put drivers in a situation in which they do not feel in danger so that the parasympathetic nervous system is predominantly active. Drivers potentially feel more comfortable just having to talk than being asked to perform roadside impairment tests. At the point where sympathetic activity switches to parasympathetic activity, impairments are expected to become more noticeable [49]. However, there are also drugs that act on the parasympathetic nervous system. Therefore, it is important to keep in mind that it is more a matter of identifying strong deviations between the behavior at the beginning when the sympathetic nervous system is highly active and the behavior during the control when the driver is in a more relaxed condition.

References

1. WHO. *Global Status Report on Road Safety*; WHO: Geneva, Switzerland, 2018.
2. García-Herrero, S.; Gutiérrez, J.M.; Herrera, S.; Azimian, A.; Mariscal, M.A. Sensitivity Analysis of Driver's Behavior and Psychophysical Conditions. *Saf. Sci.* **2020**, *125*, 104586. [[CrossRef](#)]
3. Vuurman, E.F.P.M.; Vuurman, L.L.; Lutgens, I.; Kremer, B. Allergic Rhinitis Is a Risk Factor for Traffic Safety. *Allergy* **2014**, *69*, 906–912. [[CrossRef](#)] [[PubMed](#)]
4. Orriols, L.; Delorme, B.; Gadegbeku, B.; Tricotel, A.; Contrand, B.; Laumon, B.; Salmi, L.-R.; Lagarde, E. Prescription Medicines and the Risk of Road Traffic Crashes: A French Registry-Based Study. *PLoS Med.* **2010**, *7*, e1000366. [[CrossRef](#)] [[PubMed](#)]

5. Bragazzi, N.L.; Dini, G.; Toletone, A.; Rahmani, A.; Montecucco, A.; Massa, E.; Manca, A.; Guglielmi, O.; Garbarino, S.; Debarbieri, N.; et al. Patterns of Harmful Alcohol Consumption among Truck Drivers: Implications for Occupational Health and Work Safety from a Systematic Review and Meta-Analysis. *Int. J. Environ. Res. Public Health* **2018**, *15*, 1121. [[CrossRef](#)] [[PubMed](#)]
6. Verster, J.C.; Mets, M.A.J. Psychoactive Medication and Traffic Safety. *Int. J. Environ. Res. Public Health* **2009**, *6*, 1041–1054. [[CrossRef](#)]
7. Siliquini, R.; Piat, S.C.; Alonso, F.; Druart, A.; Kedzia, M.; Mollica, A.; Siliquini, V.; Vankov, D.; Villerusa, A.; Manzoli, L. A European Study on Alcohol and Drug Use among Young Drivers: The TEND by Night Study Design and Methodology. *BMC Public Health* **2010**, *10*, 205. [[CrossRef](#)]
8. Siliquini, R.; Bert, F.; Alonso, F.; Berchiolla, P.; Colombo, A.; Druart, A.; Kedzia, M.; Siliquini, V.; Vankov, D.; Villerusa, A.; et al. Correlation between Driving-Related Skill and Alcohol Use in Young-Adults from Six European Countries: The TEN-D by Night Project. *BMC Public Health* **2011**, *11*, 526. [[CrossRef](#)]
9. Rehm, J.; Mathers, C.; Popova, S.; Thavorncharoensap, M.; Teerawattananon, Y.; Patra, J. Global Burden of Disease and Injury and Economic Cost Attributable to Alcohol Use and Alcohol-Use Disorders. *Lancet* **2009**, *373*, 2223–2233. [[CrossRef](#)]
10. Peacock, A.; Leung, J.; Larney, S.; Colledge, S.; Hickman, M.; Rehm, J.; Giovino, G.A.; West, R.; Hall, W.; Griffiths, P.; et al. Global Statistics on Alcohol, Tobacco and Illicit Drug Use: 2017 Status Report. *Addiction* **2018**, *113*, 1905–1926. [[CrossRef](#)]
11. Bush, D.M. The U.S. Mandatory Guidelines for Federal Workplace Drug Testing Programs: Current Status and Future Considerations. *Forensic Sci. Int.* **2008**, *174*, 111–119. [[CrossRef](#)]
12. Agius, R.; Kintz, P. Guidelines for European Workplace Drug and Alcohol Testing in Hair. *Drug Test. Anal.* **2010**, *2*, 367–376. [[CrossRef](#)] [[PubMed](#)]
13. Treglia, M.; Pallocci, M.; Ricciardi-Tenore, G.; Baretta, F.; Bianco, G.; Castellani, P.; Pizzuti, F.; Ottaviano, V.; Passalacqua, P.; Leonardi, C.; et al. Policies and Toxicological Screenings for No Drug Addiction: An Example from the Civil Aviation Workforce. *Int. J. Environ. Res. Public Health* **2022**, *19*, 1501. [[CrossRef](#)]
14. IRM. *Institut Für Rechtsmedizin: Jahresbericht 2019*; IRM: Zurich, Switzerland, 2020.
15. Arkell, T.R.; Kevin Richard, C.; Stuart, J.; Lintzeris, N.; Haber, P.S.; Ramaekers, J.G.; McGregor, I.S. Detection of Δ^9 THC in Oral Fluid Following Vaporized Cannabis with Varied Cannabidiol (CBD) Content: An Evaluation of Two Point-of-Collection Testing Devices. *Drug Test. Anal.* **2019**, *11*, 1486–1497. [[CrossRef](#)] [[PubMed](#)]
16. Verstraete, A.; Raes, E. *Rosita-2 Project: Final Report*; Verstraete, A.G., Raes, E., Eds.; Ghent University: Gent, Belgium, 2006.
17. Talpins, S.; Hayes, C.; Kimball, T. *The Drug Evaluation and Classification (DEC) Program: Saving Lives and Preventing Crashes*; The Traffic Law Center: Arlington, VA, USA, 2018.
18. Zielke, M.A.; Lenox, S.; Magallanes-Duarte, L.-A.; Marquart, C.; Dusek, M.; Lowe, K. Enhancing Horizontal Gaze Nystagmus Test Training through Virtual Human Simulation. *Police Pract. Res.* **2022**, 1–20. [[CrossRef](#)]
19. Boll, J. Erkennen von Fahrunfähigkeit Mittels “Verify“-Verfahren Und Beweiserbringung. In *Jahrbuch zum Strassenverkehrsrecht*; Dike Verlag: Zürich/St. Gallen, Switzerland, 2019; pp. 217–224.
20. Hahm, S.; Knuth, D.; Kehl, D.; Schmidt, S. The Impact of Different Natures of Experience on Risk Perception Regarding Fire-Related Incidents: A Comparison of Firefighters and Emergency Survivors Using Cross-National Data. *Saf. Sci.* **2016**, *82*, 274–282. [[CrossRef](#)]
21. Xu, X.; Chang, C.C.; Faber, G.S.; Kingma, I.; Dennerlein, J.T. The Validity and Interrater Reliability of Video-Based Posture Observation during Asymmetric Lifting Tasks. *Hum. Factors* **2011**, *53*, 371–382. [[CrossRef](#)]
22. Creaton, A.; Liew, D.; Knott, J.; Wright, M. Interrater Reliability of the Australasian Triage Scale for Mental Health Patients. *Emerg. Med. Australas.* **2008**, *20*, 468–474. [[CrossRef](#)] [[PubMed](#)]
23. Kobak, K.A.; Williams, J.B.W.; Engelhardt, N. A Comparison of Face-to-Face and Remote Assessment of Inter-Rater Reliability on the Hamilton Depression Rating Scale via Videoconferencing. *Psychiatry Res.* **2008**, *158*, 99–103. [[CrossRef](#)]
24. Hartwig, M.; Granhag, P.A.; Strömwall, L.A.; Vrij, A. Police Officers’ Lie Detection Accuracy: Interrogating Freely Versus Observing Video. *Police Q.* **2004**, *7*, 429–456. [[CrossRef](#)]
25. Paquette, E.; Prince, J. The Effectiveness of Simulation in Drug Recognition Expert Training: Quebec Hybrid Training Model. *Police Pract. Res.* **2020**, *22*, 510–521. [[CrossRef](#)]
26. Barrows, H.S. An Overview of the Uses of Standardized Patients for Teaching and Evaluating Clinical Skills. *AAMC. Acad. Med.* **1993**, *68*, 443–451. [[CrossRef](#)] [[PubMed](#)]
27. Aust, F.; Dienenhofen, B.; Ullrich, S.; Musch, J. Seriousness Checks Are Useful to Improve Data Validity in Online Research. *Behav. Res. Methods* **2013**, *45*, 527–535. [[CrossRef](#)] [[PubMed](#)]
28. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed.; Routledge: Hillsdale, NJ, USA, 1988.
29. Fleiss, J.L.; Levin, B.; Paik, M.C. *Statistical Methods for Rates and Proportions*, 3rd ed.; Wiley Series in Probability and Statistics; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2003; ISBN 0471526290.
30. Landis, J.R.; Koch, G.G. The Measurement of Observer Agreement for Categorical Data. *Biometrics* **1977**, *33*, 159. [[CrossRef](#)] [[PubMed](#)]
31. Olson, D.M.; Stutzman, S.; Saju, C.; Wilson, M.; Zhao, W.; Aiyagari, V. Interrater Reliability of Pupillary Assessments. *Neurocrit. Care* **2016**, *24*, 251–257. [[CrossRef](#)]
32. Angkaw, A.C.; Tran, G.Q.; Haaga, D.A.F. Effects of Training Intensity on Observers’ Ratings of Anxiety, Social Skills, and Alcohol-Specific Coping Skills. *Behav. Res. Ther.* **2006**, *44*, 533–544. [[CrossRef](#)]

33. Al-Mekhlafi, A.B.A.; Isha, A.S.N.; Chileshe, N.; Abdulrab, M.; Saeed, A.A.H.; Kineber, A.F. Modelling the Relationship between the Nature of Work Factors and Driving Performance Mediating by Role of Fatigue. *Int. J. Environ. Res. Public Health* **2021**, *18*, 6752. [[CrossRef](#)]
34. Rodríguez-Garzón, I.; Martínez-Fiestas, M.; Darmohraj, A.; Delgado-Padial, A.; Chumpitaz, R. Voluntary and Involuntary Risk Acceptance: A Case Study of Firefighters. *Saf. Sci.* **2021**, *142*, 105394. [[CrossRef](#)]
35. Hall, J.A.; Gunnery, S.D.; Horgan, T.G. Gender Differences in Interpersonal Accuracy. In *The Social Psychology of Perceiving Others Accurately*; Hall, J.A., Schmid Mast, M., West, T.V., Eds.; Cambridge University Press: Cambridge, UK, 2016; pp. 309–327.
36. Blanch-Hartigan, D.; Andrzejewski, S.A.; Hill, K.M. Training People to Be Interpersonally Accurate. In *The Social Psychology of Perceiving Others Accurately*; Hall, J.A., Schmid Mast, M., West, T.V., Eds.; Cambridge University Press: Cambridge, UK, 2016; pp. 253–269. ISBN 9781107499072.
37. Ihlebæk, C.; Løve, T.; Eilertsen, D.E.; Magnussen, S. Memory for a Staged Criminal Event Witnessed Live and on Video. *Memory* **2010**, *11*, 319–327. [[CrossRef](#)]
38. Dagar, M.; Yadav, S.; Sai, V.V.R.; Satija, J.; Bhatia, H. Emerging Trends in Point-of-Care Sensors for Illicit Drugs Analysis. *Talanta* **2022**, *238*, 123048. [[CrossRef](#)]
39. Palamar, J.J.; Ciccarone, D.; Rutherford, C.; Keyes, K.M.; Carr, T.H.; Cottler, L.B. Trends in Seizures of Powders and Pills Containing Illicit Fentanyl in the United States, 2018 through 2021. *Drug Alcohol Depend.* **2022**, *234*, 109398. [[CrossRef](#)]
40. Kumar, T.K.V. Variation in the Perception of Desired Qualities of Police Officers among Trainees and Senior Police Officers. Insights into the Process and Efficacy of Police Training. *Int. J. Comp. Appl. Crim. Justice* **2018**, *43*, 241–262. [[CrossRef](#)]
41. EMCDDA. *European Drug Report 2021: Trends and Developments*; EMCDDA: Luxembourg, 2021.
42. Lakämper, S.; Keller, K. Sleep and the Fitness to Drive: A Swiss Perspective. *Sleep Med. Evol. Contemp. Sleep Pharmacother.* **2021**. [[CrossRef](#)]
43. Ott, B.R.; Davis, J.D.; Papandonatos, G.D.; Hewitt, S.; Festa, E.K.; Heindel, W.C.; Snellgrove, C.A.; Carr, D.B. Assessment of Driving-Related Skills Prediction of Unsafe Driving in Older Adults in the Office Setting. *J. Am. Geriatr. Soc.* **2013**, *61*, 1164–1169. [[CrossRef](#)]
44. Wall, I.F.; Karch, S.B. Traffic Medicine. In *Clinical Forensic Medicine*; Stark, M.M., Ed.; Humana Press: Totowa, NJ, USA, 2005; pp. 351–385. ISBN 978-1-58829-368-8.
45. Wickens, C.D.; Carswell, C.M. Information Processing. In *Handbook of Human Factors and Ergonomics*, 4th ed.; Wiley: Hoboken, NJ, USA, 2012; pp. 117–161. [[CrossRef](#)]
46. Young, K.L.; Osborne, R.; Grzebieta, R.; Williamson, A.; Haworth, N.; Senserrick, T.; Stephan, K.L.; Charlton, J.L. Using Naturalistic Driving Data to Examine How Drivers Share Attention When Engaging in Secondary Tasks. *Saf. Sci.* **2020**, *129*, 104841. [[CrossRef](#)]
47. Dawson, D.; Reid, K. Fatigue, Alcohol and Performance Impairment. *Nature* **1997**, *388*, 235. [[CrossRef](#)]
48. Myers, D.G. *Psychologie*; Springer-Lehrbuch; Springer: Berlin/Heidelberg, Germany, 2014; ISBN 978-3-642-40781-9.
49. Ziegler, M.G. Psychological Stress and the Autonomic Nervous System. In *Primer on the Autonomic Nervous System*; Robertson, D., Biaggioni, I., Burnstock, G., Low, P.A., Eds.; Elsevier: San Diego, CA, USA, 2004; pp. 189–190. ISBN 9780080473963.