

Validation of the Clear and Unambiguous Human Machine Interface Questionnaire

Marijke van Weperen¹, Susanne de Jong¹, Mervyn Franssen², Hari Subraveti¹,
Marieke Martens^{1,2}, & Jeroen Hogema¹
¹TNO Integrated Vehicle Safety,
²Eindhoven University of Technology,
Netherlands

Abstract

With the ongoing development of Advanced Driver Assistance Systems and Automated Driving Systems, driving is becoming increasingly automated. This causes a growing importance of ensuring safe interaction between drivers and systems with different levels of automation. A key requirement for information presented to users through the Human Machine Interface is that it must be “clear and unambiguous”. In previous research, we designed a questionnaire capable of distinguishing between HMIs with varying levels of clarity and unambiguity. The present study extends this work by providing additional validation in a more operational setting. Thirty-seven participants evaluated 24 different HMIs in a driving simulator using the questionnaire. The HMIs were designed to be good, average or bad in terms of clarity and unambiguity. Two groups of participants were selected: either with or without experience with Advanced Driver Assistance Systems. This validation builds on earlier research by incorporating more representative HMI designs across various quality levels and user groups. The results suggest that the questionnaire discriminated among the different designs for both user groups, indicating that the questionnaire is a valid tool for evaluating HMI clarity and unambiguity, independent of the user’s prior experience with automation.

Introduction

As Advanced Driver Assistance Systems (ADAS) and Automated Driving Systems (ADS) continue to advance, more automation levels are becoming available on production vehicles. Now and for the foreseeable future, most automation remains partial or conditional, requiring drivers to be aware of what the system is doing and when and how to take back control. Effective interaction between the driver and the system, potentially consisting of different sets of automated functionalities depending on conditions, is key to avoid mode confusion. Requirements for this interaction are currently available only in broad and abstract terms. One fundamental requirement for the system information presented to the driver via the Human Machine Interface (HMI) is that it needs to be easy to understand. Guidelines and Recommendations concerning Safety Requirements for Automated Driving Systems (GRVA, 2022) emphasize that the information provided by the HMI to the user needs to be “clear

and unambiguous”. In order to evaluate the HMI effectively, it is necessary to have a standardized assessment tool for these terms.

In a previous study (Hogema et al., 2024), the terms “clear and unambiguous” were operationalized in the Clear and Unambiguous Questionnaire (CUQ). The CUQ covers core aspects of clarity and unambiguity by focussing on users’ comprehension and perception. While both the widely used User Experience Questionnaire (UEQ) (Laugwitz et al., 2008) and the System Usability Scale (SUS) (Brooke, 1995) contain items that touch upon these factors, they measure usability of the design in a broader sense. Other questionnaires, such as the Van der Laan System Acceptance Scale (Van der Laan et al., 1997) and the Usability Metric for User Experience (UMUX, Finstad, 2010) can be used to compare HMI designs in terms of usability and acceptance, but do not specifically assess clarity and unambiguity of the interface either.

The CUQ has previously been evaluated in a desktop environment where it has been shown to differentiate between good and bad HMI designs in terms of clarity and unambiguity. In this paper, we assess the validity and reliability of the CUQ in a driving simulator environment, which is a significant step closer to its application in a real vehicle. The current validation of the questionnaire focusses on its ability to distinguish among good, average and bad quality HMI designs in terms of clarity and unambiguity, expanding previous work where only good and bad designs were distinguished. Similarly to previous work, the questionnaire is validated using different designs and automation levels. HMIs under evaluation cover various SAE levels of automation (SAE, 2022): manual driving, partial automation (Level 2, L2: hands-on and hands off) and conditional automation (Level 3, L3). We also investigate whether CUQ responses vary based on people’s experience with ADAS.

Method

Participants

Participants qualified for the study if they had held a driving licence for at least five years and drove a minimum of 5000 km annually. Participants were categorized as either experienced or inexperienced with L2 automation systems based on self-reporting. Participants who had driven vehicles equipped with Adaptive Cruise Control (ACC) and Lane Centring Assist (LCA) at least once per month during the past year were considered experienced. The study involved 37 participants in total: 17 (14 males and 3 females) had experience with the L2 system, while the other 20 (9 males, 10 females, and 1 non-binary) did not. Distributions of age and driving experience are presented in Table 1. All participants provided written informed consent before participating and were paid €30 for their contribution. The study received approval from our internal review board for studies involving human participants (Nr. 2025-671).

Table 1. Participant age and driving experience distributions by user group.

<i>Age (years)</i>	<i>Experienced</i>	<i>Inexperienced</i>
23-30	2	1
31-40	4	4
41-50	4	4
51-67	7	11
<i>Driving experience (1000km/year)</i>		
5-10	0	3
11-20	4	9
21-30	7	4
31-40	1	3
> 40	5	1

Apparatus

The experiment was conducted in a driving simulator to create a visually immersive environment to assess the HMIs. The driving simulator consisted of a BMW body on a motion platform and a 4m high and 230° cylindrical screen on which three projectors presented the dynamical virtual environment. The motion platform was stationary during this experiment to purely assess the HMI. The digital cockpit of the driving simulator allowed to show any kind of dashboard (dimensions of 243x90 mm, resolution of 1920x720 pixels) to the participant. LEDs were integrated in the rim of the steering wheel and an audio system provided participants with vehicle sounds and HMI warning sounds.

Study design

Participants experienced 24 scenarios presented in random order, summarised in Table 2. In these scenarios, the vehicle drove at a constant speed of 100 km/h on a three-lane motorway, on a different lane per scenario for visual variety. Traffic was light across all lanes, with vehicles driving at 105 km/h in the left lane, 100 km/h in the middle lane and 95 km/h in the right lane. Traffic was kept constant across scenarios. Participants were seated in the driver's seat of the simulator, experiencing the drive from the driver's perspective. Participants were instructed that their only task was to judge the HMIs and not to drive or perform any interaction with the systems. To cover a wide range of HMI designs, the scenarios differed in three main aspects. First, the scenarios varied by type: scenarios were either static (i.e., the HMI info remained unchanged) or involved a transition from a level of automated (LoA) to manual control. Second, the scenarios differed in the LoA, ranging from manual driving to conditional automated driving. A combination of scenario Type and LoA is referred to as a scenario Version. Third, each version featured a bad, average, and good HMI design.

Table 2. Experiment scenarios (3x8=24 in total; the numbers represent the scenario number).

Type	Level of automation	Version	HMI quality		
			Bad	Average	Good
Static	L0	A	1	2	3
	L2 hands-on	B	4	5	6
	L2 hands-off	C	7	8	9
	L3	D	10	11	12
Transition	L2 hands-on to L0	E	13	14	15
	L3 to L0 (version 1)	F	16	17	18
	L3 to L0 (version 2)	G	19	20	21
	L2 hands-off to L0	H	22	23	24

HMI designs

The HMI designs were inspired by current production vehicle HMIs and relevant literature, e.g., Perrier et al. (2021). A team of four experts in vehicle automation and user interaction used various design factors from Table 3 to design good, average, and bad HMI designs in terms of clarity and ambiguity. Typically, clear HMI designs consisted of text-based driver instructions and intuitive function naming. Unintuitive function naming, a larger amount of (ambiguous) symbols and symbols with orange and red colours led to average or bad designs in terms of clarity and ambiguity. Table 4 shows all static HMI versions for the three quality levels. Font sizes and contrast were not varied across designs, to prevent HMIs from being evaluated badly due to readability limitations rather than design differences.

Table 3. HMI design factors.

Design factor	Description
Amount of information	The amount of information HMIs provided to the driver.
Sizes of symbols	Both smaller and bigger symbols.
Variety of symbols	Different symbols for assistance systems.
Colours of symbols	Different colours for activation status.
Positioning of symbols and text	Variation in placement of symbols and text.
Own vehicle visualisation	Variation in own vehicle visualization, showing the driver what has been perceived by the vehicle in terms of other road users and road-layout.

The HMI designs included a digital dashboard, auditory cues, and coloured LED segments on the steering wheel. Figure 2 shows the experimental setup in the driving simulator and an HMI example. A detailed description and visualisation of each scenario is provided in Hogema et al. (2025). Each scenario lasted 40 seconds. A scenario with a transition included three HMI stages: it started with the initial state (15 s), followed by a transition (10 s), and then showed the final state (15 s). During the transition, various elements changed depending on the HMI design: dashboard symbols and text either appeared, disappeared, flashed, or changed colour and/or steering wheel LEDs blinked. In case of a transition from L3 or L2-hands-off to L0, a voice-over "You decide to grab the steering wheel" was played to explicitly indicate

that the shown HMI transition happened as if the driver grabbed the steering wheel. Without this clarification, HMIs under real-world conditions would escalate warnings, rather than moving to the intended final state defined in the study.



Figure 2. Experimental setup with an example of a good quality HMI designs in terms of clarity and unambiguity.

Table 4. Overview of the static HMI designs by level of automation and quality level.

Level of automation	Quality: good	Quality: average	Quality: bad
L0			
L2 hands-on			
L2 hands-off			
L3			

Measures

Participants answered questions (in Dutch) related to the scenarios they had just observed. The first part contained four responsibility questions (see Table 5) concerning participants' understanding of actions they would need to perform at the scenario's final state. For transition scenarios, a fifth question assessed participants' interpretation regarding the change in responsibility. The responsibility questions were included to determine whether participants who rated an HMI highly on the CUQ actually also understood their responsibilities better.

The second questionnaire was the CUQ (see Table 6). Participants responded to nine items on a 7-point scale ranging from strongly disagree to strongly agree, which assessed how clear and unambiguous they found the information presented by each HMI. The questionnaire was slightly adapted from the previous study by inverting items 3, 5, 7, and 8. This resulted in a similar amount of positive and negative statements belonging to each dimension

Table 5. Responsibility questions (translated to English; Item 5 was only asked for transition scenarios).

Item	Question	Response options
1	Who needs to accelerate and brake in this state?	(A) I do this all by myself; (B) The car, but I constantly need to be able to intervene; (C) The car does this all by itself; (D) I don't know
2	Who needs to steer in this state?	(A) I do this all by myself; (B) The car, but I constantly need to be able to intervene; (C) The car does this all by itself; (D) I don't know
3	Are you obligated to keep an eye on the environment during driving in this state?	(A) Yes; (B) No; (C) I don't know
4	Are you obligated to keep your hands on the steering wheel in this state?	(A) Yes; (B) No; (C) I don't know
5	Which of the following tasks changed during this transition? Multiple answers possible	(A) Accelerating/Braking; (B) Steering; (C) Keeping an eye on the environment; (D) Keeping hands on the steering wheel; (E) I don't know

Table 6. *Clear and Unambiguous Questionnaire (translated to English).*

Question item	Question
1	The information is ambiguous
2	This information can have multiple meanings
3	I am certain what this information means
4	This information is confusing
5	This information is clear
6	I can quickly comprehend this information
7	The information is illegible
8	The information is hard to perceive
9	The information is presented in a clear manner

Procedure

At the start of the experiment, participants were informed about the experimental setup, their tasks and provided informed consent. To familiarize participants with the voice-over, they listened to an example of the voice-over. Next, participants completed a questionnaire about their demographics and driving experience.

Participants were introduced to the experimental setup through two practice scenarios. Subsequently, all 24 scenarios were presented. The transition scenarios were played twice in a row to give participants sufficient opportunity to observe what happened during the scenario. To reduce memory burden, the final state of the HMI remained visible on the dashboard while participants answered the questions. The questions were answered immediately after each scenario using a tablet mounted on the centre console. After half of the scenarios were completed, participants were given a break of approximately 10 minutes before proceeding with the remaining scenarios.

Once all scenarios were finished, participants took part in a semi-structured interview to discuss their experiences. This paper does not discuss the results of this interview, as the primary focus is on the validation of the CUQ. The entire procedure took about 2.5 hours for each participant.

Data analysis

The raw CUQ scores ranging from 1 to 7 were mapped to a range from -3 to 3 and inverted if needed, such that positive values indicated a favourable HMI assessment. An exploratory factor analysis, using maximum likelihood with oblique rotation (direct oblimin), was performed on the nine items. The number of factors was determined using the scree plot and the elbow criterion, in combination with the Kaiser criterion, and factor scores were calculated as weighted averages based on their loadings. The resulting scores “perception” and “comprehension” were analysed using a linear mixed-effects model (LMM), with the formula:

$$\text{Score} \sim \text{Quality} * \text{Experience} + (1|\text{Participant}) + (1|\text{Version}).$$

Quality and Experience were entered as categorical predictors using treatment (dummy) coding. For Quality, the reference category was the average quality, and for

Experience, the reference category was no experience. Type III ANOVA assessed fixed effect differences, and significant results were followed up with Tukey-corrected post hoc *comparisons of the estimated marginal means*.

Answers to the responsibility questions were evaluated as either correct or incorrect (answers of "I don't know" were considered incorrect). The correctness of the responsibility questions was analysed using a binomial generalised LMM (GLMM), specified as:

$$\begin{aligned} & \text{Correctness} \sim (\text{Comprehension score} + \text{Perception score}) * \\ & \text{Experience} + (\text{Comprehension score} + \text{Perception score}) * \\ & \text{Responsibility} + (1|\text{Participant}) + (1|\text{Version}). \end{aligned}$$

Experience and Responsibility were entered as categorical predictors using treatment (dummy) coding. The reference category for Experience was "no experience", while the reference category for Responsibility was item 1 of the responsibility questions. Including responsibility as a factor allowed to assess whether high perception and comprehension scores corresponded to correct responsibility answers for each driving responsibility (i.e., who must accelerate, brake, and steer, and whether the driver is required to monitor the environment and keep their hands on the steering wheel). Item 5 of the responsibility questions was excluded from the analysis because it was asked only in transition scenarios. The model also included random intercepts for participant and scenario version.

LMMs were conducted in R (v4.4.2) using *lmerTest* (v3.1.3, with REML estimation), and the GLMM with *lme4* (v1.1.37). The *p*-value of the fixed effects were calculated using Satterthwaite approximation and Type III Sum of Squares. Confidence intervals were calculated with the Wald method. Results were considered significant at $p < .05$.

Results

Clear and Unambiguous Questionnaire

Figure 3 shows participant responses to each item of the CUQ across all scenarios, separated by the HMI quality. For most items, the graphs indicate that participants rated good HMIs more positively and bad HMIs more negatively. The exceptions are items 7 (difficult to read) and 8 (difficult to perceive), as over 90% of participants indicated disagreement or strong disagreement, irrespective of the HMI quality. The CUQ showed a high internal consistency (Cronbach's $\alpha = 0.92$, 95% CI: [0.916, 0.931]), and the data sample was suitable for analysis (KMO = .91). Two factors were extracted following the scree plot (elbow criterion) and the Kaiser criterion, accounting for 79.4% of variance. Table 7 shows the factor loadings.

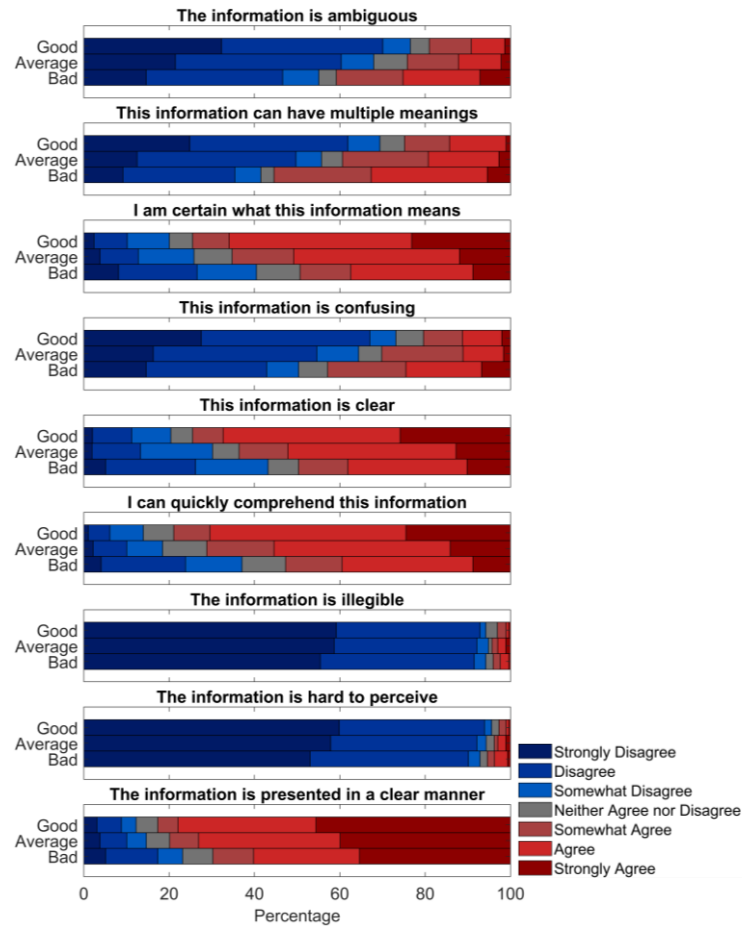


Figure 3. CUQ responses on the 7-point scale for each item across all scenarios, categorized by the HMI quality levels: bad, average and good.

Table 7. Pattern matrix with factor loadings when extracting two factors ranked highest to lowest (values < 0.3 are suppressed; * are negatively phrased questions that were inverted before the factor analysis).

Item	Factor 1	Factor 2
3 I am certain what this information means	0.93	
5 This information is clear	0.93	
4 This information is confusing *	0.91	
1 The information is ambiguous *	0.87	
2 This information can have multiple meanings *	0.86	
6 I can quickly comprehend this information	0.86	
9 The information is presented in a clear manner	0.51	0.31
8 The information is hard to perceive *		0.91
7 The information is illegible *		0.87

Items grouped on the same factor were labelled as comprehension (Factor 1) and perception (Factor 2). Item 9, “The information is presented in a clear manner,” related to both factors. Cronbach’s alpha was 0.95 for comprehension and 0.72 for perception. Removing item 9 increased perception’s alpha to 0.86, showing item 9 only weakened this factor. Therefore, only items 7 and 8 were used for perception scores. Comprehension and perception scores were calculated as weighted averages based on the factor loadings in Table 7. Over 90% of perception scores were between 2 and 3, while comprehension ranged from -3 to 3 (clustering around 2). Comprehension median scores were 0.50, 1.68, and 2.00 (*SDs* 1.70, 1.48, 1.44) and perception medians were 3.00, 3.00, and 3.00 (*SDs* 1.02, 1.02, 0.89) for bad, average, and good HMI qualities, respectively. Tables 8 and 9 show LMM results for comprehension ($R_m^2 = 0.08$, $R_c^2 = 0.40$) and perception ($R_m^2 = 0.02$, $R_c^2 = 0.44$), and Tables 10 and 11 present the type III ANOVA results of the LMMs.

Table 8. Linear Mixed Model results of the comprehension score.

	<i>Estimate</i>	<i>SE</i>	<i>95% CI</i>	<i>t</i>	<i>p</i>
Intercept	1.12	0.24	[0.65, 1.59]	4.65	< .001
Quality: bad	-0.68	0.14	[-0.96, -0.41]	-4.83	< .001
Quality: good	0.43	0.14	[0.15, 0.71]	3.05	.002
Experience: yes	-0.09	0.33	[-0.73, 0.54]	-0.29	.774
Quality: bad *	-0.13	0.21	[-0.54, 0.28]	-0.62	.537
Experience: yes					
Quality: good *	-0.22	0.21	[-0.63, 0.19]	-1.05	.294
Experience: yes					
	<i>Variance</i>	<i>SD</i>	<i>Number of groups</i>		
Participant (Intercept)	0.78	0.88	37		
Version (Intercept)	0.07	0.27	8		

Table 9. Linear Mixed Model results of the perception score.

	Estimate	SE	95% CI	t	p
Intercept	2.55	0.15	[2.25, 2.85]	16.54	< .001
Quality: bad	-0.07	0.08	[-0.24, 0.09]	-0.90	.369
Quality: good	-0.06	0.08	[-0.22, 0.10]	-0.72	.473
Experience: yes	-0.26	0.23	[-0.71, 0.18]	-1.16	.251
Quality: bad *	-0.11	0.12	[-0.35, 0.12]	-0.94	.348
Experience: yes					
Quality: good *	0.15	0.12	[-0.09, 0.39]	1.25	.212
Experience: yes					
	Variance	SD	Number of groups		
Participant (Intercept)	0.403	0.63	37		
Version (Intercept)	0.001	0.04	8		

Table 10. Type III ANOVA on the comprehension score LMM.

	SS	MS	NumDF	DenDF	F	p
Quality	175.24	87.62	2	834.04	55.54	< .001
Experience	0.76	0.76	1	35.01	0.48	.493
Quality * Experience	1.76	0.88	2	834.04	0.56	.572

Table 11. Type III ANOVA on the perception score LMM.

	SS	MS	NumDF	DenDF	F	p
Quality	3.85	1.92	2	834.04	3.55	.029
Experience	0.74	0.74	1	35.01	1.36	.252
Quality * Experience	2.61	1.31	2	834.04	2.41	.090

There was an effect of HMI quality on both comprehension and perception scores. Tukey-corrected post-hoc pairwise comparisons of the estimated marginal means show comprehension score was lower for bad ($EMM = 0.33$, 95% CI [-0.06, 0.71]) compared to average ($EMM = 1.07$, 95% CI [0.69, 1.46]) quality HMIs ($t(834) = -7.18$, $p < .001$) and bad to good ($EMM = 1.39$, 95% CI [1.01, 1.78]) quality HMIs ($t(834) = -10.27$, $p < .001$). The average quality scored also lower than the good quality HMIs ($t(834) = -3.09$, $p = .006$). For perception score, a significant difference was found only between bad ($EMM = 2.29$, 95% CI [2.06, 2.52]) and good ($EMM = 2.43$, 95% CI [2.20, 2.66]) quality HMIs ($t(834) = -2.43$, $p = .041$). No significant differences were found between average ($EMM = 2.42$, 95% CI [2.19, 2.65]) and bad ($t(834) = 2.16$, $p = .080$), or between average and good ($t(834) = -0.27$, $p = .960$). Experience had no significant effect on either score.

Responsibility questions

Table 12 presents the results of the binomial GLMM testing whether correctness on responsibility questions depended on comprehension and perception scores ($R_m^2 = 0.12$, $R_c^2 = 0.23$). Comprehension score significantly predicted correctness on the responsibility questions, with higher comprehension scores associated with greater

odds of a correct response ($\beta = 0.19$). This effect varied across responsibility items. For the environment monitoring question (item 3), comprehension score had little effect on correctness, reflected by a significant negative interaction ($\beta = -0.01$). In contrast, the positive relationship between comprehension score and correctness was stronger for the question related to hands on the steering wheel (item 4), reflected by a significant positive interaction ($\beta = 0.34$). Participants were also more likely to answer items 3 and 4 correctly than the reference responsibility item 1 (acceleration question).

Table 12. Generalised Linear Mixed Model results of the correctness of the responsibly questions.

	Estimate	SE	z	p	OR	OR 95% CI
Intercept	0.34	0.33	1.06	.290	1.41	[0.74, 2.68]
Comprehension score	0.19	0.06	2.20	.001	1.21	[1.08, 1.36]
Perception score	-0.02	0.10	-0.17	.867	0.98	[0.81, 1.20]
Experience: yes	0.16	0.31	0.52	.603	1.17	[0.64, 2.14]
Responsibility: item 2	0.22	0.28	0.79	.430	1.25	[0.72, 2.15]
Responsibility: item 3	1.29	0.31	4.15	< .001	3.64	[1.98, 6.71]
Responsibility: item 4	0.79	0.29	2.75	.006	2.21	[1.26, 3.90]
Comprehension score * Experience: yes	-0.01	0.06	-0.18	.860	0.99	[0.87, 1.12]
Perception score * Experience: yes	-0.05	0.11	-0.46	.647	0.95	[0.76, 1.18]
Comprehension score * Responsibility: item 2	0.04	0.07	0.54	.592	1.04	[0.90, 1.19]
Comprehension score * Responsibility: item 3	-0.20	0.08	-2.43	.015	0.82	[0.69, 0.96]
Comprehension score * Responsibility: item 4	0.15	0.08	2.01	.044	1.17	[1.00, 1.35]
Perception score * Responsibility: item 2	0.12	0.11	1.02	.306	1.12	[0.90, 1.41]
Perception score * Responsibility: item 3	0.19	0.13	1.46	.143	1.21	[0.94, 1.56]
Perception score * Responsibility: item 4	0.13	0.12	1.07	.286	1.14	[0.90, 1.43]
	Variance	SD	Number of groups			
Participant (Intercept)	0.23	0.48	37			
Version (Intercept)	0.25	0.50	8			

Discussion and conclusion

Similar to Hogema et al. (2024), the questionnaire showed good internal validity and the factor analysis identified the same two underlying factors, namely comprehension and perception. An LMM analysis showed that there was an influence of quality on both comprehension and perception scores. The comprehension scores differed significantly between good, average and bad quality HMI designs. Experience did not have a significant influence on these two factors, suggesting that the questionnaire can be used by both experienced L2 drivers and inexperienced drivers.

Additionally, analysis of the responsibility questions revealed a generally positive effect of HMI comprehension on the correctness of responsibility questions. One exception to this were the responsibility questions about monitoring the environment, for which no effect was identified. This means that when people rated an HMI with a higher comprehension score, they generally also demonstrated a better understanding of their driving responsibilities. The authors advise to include the responsibility questions in the methodology of HMI assessment, since the actual understanding of the drivers' role is a safety concern. The worst HMI is the one which drivers think they understand, while their understanding is incorrect.

As already outlined in the previous study, safe interaction relies on much more factors than only perception and comprehension. Perception and comprehension of the HMI design is a first prerequisite, but are insufficient to guarantee safe interaction. Additional objective measures, such as response times and type of response (braking or steering), are necessary to ensure that the drivers' response to system cues are both timely and correct.

Limitations

Several limitations of this study should be acknowledged. The sample size of this study does not allow for a confirmatory analysis of the two-factor model, therefore only an exploratory analysis was performed. Sample size may also have obscured the impact of experience. Furthermore, participants with different levels of experience might have differed on other measured or unmeasured characteristics as well, which could potentially have masked the effect on experience.

Perception scores only differed between good and bad quality HMI designs, likely due to the design choice that many elements related to perception, i.e., font size and contrast, remained constant over all HMI designs. In future studies, it is advised to develop various benchmark HMIs which function as a reference for the questionnaire results, such that all items of the questionnaire are validated.

In this study, the HMI was evaluated under conditions in which the driver passively observed the system, without engaging in active interaction. However, it is essential to investigate if the questionnaire remains applicable in dynamic traffic environments, where the driver might have limited time to observe the HMI.

Conclusion

In conclusion, this study provides evidence for the validity of the clear and unambiguous questionnaire for different user groups in a driving simulator study with realistic HMI designs. The results indicate that the questionnaire is able to differentiate among HMI designs in terms of clarity and unambiguity and that the usability seems independent of experience with L2 systems. For further development of the questionnaire, it is advised to use benchmarked HMIs of various qualities and to use the questionnaire in dynamic traffic scenarios and conditions.

Acknowledgement

This project was financed by the Dutch Ministry of Infrastructure and Water Management.

References

- Brooke, J. (1995). SUS: A quick and dirty usability scale. In P.W. Jordan, B. Thomas, B.A. Weerdmeester, and A.L. McClelland (Eds.), *Usability Evaluation in Industry* (pp. 189 – 194). London, UK: Taylor and Francis.
- Finstad, K. (2010). The Usability Metric for User Experience. *Interacting with Computers*, 22, 323-327.
- GRVA. (2022). *Guidelines and Recommendations concerning Safety Requirements for Automated Driving Systems (Report WP.29-187-10)*. Geneva, Switzerland: United Nations Working Party on Automated, Autonomous, and Connected Vehicles.
- Hart, S.G. & Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology* (pp. 139-183). Amsterdam, Netherlands: North-Holland.
- Hogema, J., Post, R., Verstegen, R., & Souman, J. (2024). Automated Driving System HMIs: “Clear and Unambiguous”?, In *Proceedings of the 16th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 184 – 193). New York, USA: Association for Computing Machinery.
- Hogema, J., Van Weperen, M., Franssen, M., Subraveti, H.H.S.N., Martens, M.H., & Gaggar, S. (2025). *Clear and Unambiguous HMIs* (TNO 2025 R12106). The Hague, Netherlands: Netherlands Organization for Applied Scientific Research.
- Laugwitz, B., Held, T., & Schrepp, M. (2008). Construction and evaluation of a user experience questionnaire. In *Symposium of the Austrian HCI and usability engineering group* (pp. 63-76). Berlin, Heidelberg, Germany: Springer Berlin Heidelberg.
- Perrier, M.J.R., Louw, T.L., & Carsten, O. (2021). User-centred design evaluation of symbols for adaptive cruise control (ACC) and lane-keeping assistance (LKA). *Cognition, Technology & Work*, 23, 685-703.
- SAE. (2022). *Naming Methodology for Driving Automation Systems*. Warrendale, USA: SAE International.
- Van der Laan, J.D., Heino, A., & De Waard, D. (1997). A simple procedure for the assessment of acceptance of advanced transport telematics. *Transportation Research Part C: Emerging Technologies*, 5, 1-10.