A Framework for Disentangling Efficiency from Effectiveness in External HMI Evaluation Procedures for Automated Vehicles

Alexandros Rouchitsas[a]
Department of Information Technology, Uppsala University, Uppsala, Sweden

Keywords: External Human-Machine Interfaces, Automated Vehicles, Road Users, Usability, Evaluation, Methodology.

Abstract: Automated vehicles (AVs) are rapidly transforming smart cities, offering potential benefits such as improved safety, performance, mobility, accessibility, and overall user experience in traffic. A key area of focus in this evolution is the development of external human-machine interfaces (eHMIs) which aim to equip AVs with communication capabilities. Said interfaces address critical challenges, including mitigating safety risks and enhancing traffic flow in scenarios where drivers are inattentive or altogether absent, and play an important role in allaying distrust of the general public in AVs. Considering the research field of eHMIs is relatively young, it is unsurprising that standardized eHMI evaluation procedures are yet to be established. As a result, the effectiveness and efficiency of eHMI concepts are often assessed either simultaneously within the same evaluation procedure or separately but in otherwise similar procedures. Unfortunately, these approaches overlook on the whole the fundamental differences between the two constructs, resulting in limitations relating to the validity, reliability, and comparability of the findings. Here, I present a definitive framework aimed at disentangling efficiency from effectiveness by guiding methodological choices regarding design rationale explanation, instructions emphasizing speed, trial-level time limit, and targeted performance measures, depending on the research questions of interest.

1 INTRODUCTION

Automated vehicles (AVs) are taking smart cities by storm due to their potential for improving safety, performance, mobility, accessibility, and overall user experience in traffic. Sooner rather than later, road users will need to interact extensively and intensively with highly (SAE Level 4) and fully (SAE Level 5) automated vehicles that either transport passengers who no longer have to be attentive to the road and participate in traffic interaction scenarios or simply drive around with no human operator or passenger on board, to deliver goods and services to third parties (ISO/TR 23049:2018, 2018; SAE International J3016, 2021).

The burgeoning field of external human-machine interfaces (eHMIs) has been concerned with equipping AVs with communication capabilities to primarily mitigate traffic safety and traffic flow issues that are expected to arise due to inattentive or altogether absent drivers, but also promote public trust in and acceptance of this novel technology (Rouchitsas and Alm, 2019; Dey et al., 2020; Calvo-Barajas et al., 2025). Typical eHMI concepts provide information regarding kinematics of the oncoming vehicle (e.g., speed and acceleration), mode (e.g., manual, highly, or fully automated), situational awareness (detection and acknowledgement of nearby vulnerable road users such as pedestrians and cyclists), and imminent maneuvres (e.g., yielding, taking off, or changing lanes). To achieve this, eHMIs utilize the external surface and/or the immediate surroundings of an AV to communicate relevant messages via LED light strips, rotating headlights, displays, speakers, on-road projections, and even shape change (Bazilinskyy et al., 2019).

eHMI concepts are commonly evaluated with respect to their usability, i.e., their effectiveness, efficiency, and potential for user satisfaction, in the context of field studies, laboratory experiments, and online surveys (Rouchitsas and Alm, 2019). Effectiveness refers to a concept's ability to bring about the desired result, whereas efficiency refers to a concept's ability to be effective all the while

[a] https://orcid.org/0000-0003-3503-4676

Rouchitsas, A.
A Framework for Disentangling Efficiency from Effectiveness in External HMI Evaluation Procedures for Automated Vehicles.
DOI: 10.5220/0013435900003941

In *Proceedings of the 11th International Conference on Vehicle Technology and Intelligent Transport Systems (VEHITS 2025)*, pages 616-621
ISBN: 978-989-758-745-0; ISSN: 2184-495X

expending the least amount of resources (Bevan et al., 2015, Schömig et al., 2024). Considering this research field is relatively young, it comes as no surprise that, despite a few noteworthy attempts and calls to action, eHMI evaluation procedures have not been standardized yet (Rouchitsas and Alm, 2019; Kaß et al., 2020). As a result, the effectiveness and efficiency of eHMIs are often assessed either simultaneously within the same evaluation procedure or separately but in otherwise similar procedures. Unfortunately, these approaches overlook for the most part the fundamental differences between the two constructs, resulting in limitations relating to the validity, reliability, and comparability of the findings. In this paper, I present a definitive framework aimed at disentangling efficiency from effectiveness by guiding methodological choices regarding design rationale explanation, instructions emphasizing speed, trial-level time limit, and performance measures, depending on the research questions of interest.

In the following sections, I will delve into the proposed framework and the methodological concerns it aims to alleviate (section 2), present relevant eHMI evaluation attempts (section 3), discuss the benefits of the proposed framework (section 4), and conclude with the take-home message and suggestions for future work (section 5).

2 METHODOLOGICAL CONCERNS AND PROPOSED FRAMEWORK

The majority of empirical work has focused on evaluating the effectiveness of eHMI concepts which basically translates to communicating relevant information clearly and thus supporting road users in making appropriate decisions depending on the specifics of the traffic situation at hand. In eHMI evaluation procedures, participants are typically asked to infer what the AV is trying to communicate and decide appropriately, while their response accuracy – assessed from error rates – is treated as a measure of the effectiveness of each eHMI concept, most times referred to as the "comprehensibility", "understandability", or "intelligibility" of the concept (Bevan et al., 2015; Wiese et al., 2017). In that sense, effectiveness is evaluated on the basis of an absolute criterion: a concept can either be effective or not.

For all that, numerous studies have also evaluated eHMI concepts with respect to their efficiency, which in this case translates to communicating relevant information faster and easier than alternative concepts, and therefore leading to appropriate responses in less time and with less effort. In these eHMI evaluation procedures, participants are typically asked to make the appropriate traffic decision as fast as possible, while their response latency – assessed from reaction times (RTs) – and their effort – assessed from NASA-TLX ratings – is treated as a measure of the efficiency of each concept (Hart, 2006; Bevan et al., 2015; Wiese et al., 2017). In that sense, efficiency is evaluated on the basis of a relative criterion: a concept can fare "better" or "worse" compared to other concepts.

2.1 Effectiveness

Nevertheless, it is often the case that due to feasibility limitations or methodological confusion, effectiveness and efficiency are either evaluated simultaneously in the context of the same evaluation procedure or separately but in the context of otherwise similar procedures that do not take into account the fundamental differences between the two constructs (Cheema et al., 2023). More specifically, for effectiveness to be evaluated in a proper manner, no explanation of the design rationale behind each eHMI concept should be provided to participants beforehand to ensure unbiased responses. Moreover, no instruction to "respond as fast and accurately as possible" should be provided to ensure mitigation of the speed-accuracy trade off phenomenon, the well-known phenomenon according to which emphasizing fast responses leads to a higher percentage of incorrect ones (Kantowitz et al., 2014). Lastly, the evaluation procedure should employ self-paced trials with no time limit at the trial level to ensure participants have ample time to decode the presented message and act appropriately.

In the event that explanation is provided beforehand in evaluations of eHMI effectiveness, a *correct response* regarding "comprehensibility", "understandability", or "intelligibility" of any eHMI concept will be biased and thus rendered useless for further analysis. Accordingly, instructions to respond fast will render data uninterpretable as in the event of an *incorrect response* the simple question "Was the participant hurrying or did they truly not know the right answer?" cannot be conclusively answered. In the same vein, responding within a prespecified temporal window will render data uninterpretable in the event of *no response* as the simple question "Was the participant too slow or did they truly not know the right answer?" cannot be conclusively answered either. It becomes easily apparent then that

Table 1: Proposed framework for eHMI evaluation procedures for AVs.

		Design Rationale Explanation		**Instructions Emphasizing Speed**		**Trial-Level Time Limit**		**Performance Measures**		
		No	**Yes**	**No**	**Yes**	**No**	**Yes**	**Error Rates**	**RTs**	**NASA-TLX Ratings**
Research Question	**Effectiveness**	✔	**Correct response: Biased**	✔	**Incorrect response: “Hurried or didn’t know?”**	✔	**No response: “Too slow or didn't know?”**	✔	–	–
	Efficiency	**Correct response: Entangled**	✔	**Correct response: Indistinct**	✔	**All responses: “Too slow or not following instructions?”**	✔	✔[1]	✔	✔

effectiveness should be evaluated in the context of evaluation procedures where the design rationale is not explained to participants beforehand, instructions to respond fast are not given, and there is no time limit at the trial level.

2.2 Efficiency

Given that efficiency is meaningful only in the context of effectiveness – the first being a subpart of the second, as it is nonsensical to think of something as being efficient when it is not effective in the first place – it is reasonable to expect that the underlying psychological processes that affect both overlap to a great extent (Cheema et al., 2023). For instance, when evaluating an eHMI concept, its legibility, i.e., its quality of being clear enough to read quickly and easily, and its intelligibility interact in a way that if evaluated at the same time, a *correct response* cannot be attributed conclusively to either factor. Therefore, it is crucial to develop evaluation procedures that disentangle the two. An obvious workaround if one were to evaluate legibility would be to provide participants with the design rationale behind the eHMI concept in question beforehand, and then opt for a task where participants know in advance what the correct response is and are only evaluated on the basis of how quickly and easily they respond.

Evidently, for efficiency to be evaluated in a proper manner, emphasis should also be placed on fast responses given that it is mostly the temporal aspect that is under scrutiny in this case. Therefore, it is essential to also provide instructions to participants that emphasize speed to generate sufficiently distinct *correct responses*, given latency measures such as RTs are more sensitive to experimental manipulations and their accompanying differences (Kyllonen and Zu, 2016).

Furthermore, contrary to effectiveness evaluations, responding within a prespecified temporal window is essential in efficiency evaluations, as the absence of such a window will render data uninterpretable for *all responses* (correct; incorrect; no response) considering the simple question “Was the participant too slow or simply not following the instructions?” cannot be conclusively answered. It becomes easily apparent then that efficiency should be evaluated in the context of procedures where the design rationale is explained beforehand, instructions to respond fast are given to participants, and there is a tight time limit at the trial level.

Table 1 summarizes the optimal methodological choices for eHMI evaluations regarding design rationale explanation, emphasis on fast responses, time limit at the trial level, and targeted performance measures according to the proposed framework. It is clear that what is methodologically optimal for effectiveness evaluations should absolutely be avoided in the case of efficiency evaluations, and vice versa, if one is aiming for interpretable data.

[1] It is common practice in the experimental/cognitive psychology tradition to analyze RTs for correct responses only (Kyllonen and Zu, 2016).

3 eHMI EVALUATION ATTEMPTS

The lack of a definitive framework aimed at disentangling efficiency from effectiveness in eHMI evaluation procedures has for the most part left researchers and practitioners to their own devices and has allowed for counterproductive amounts of improvization and creative freedom to creep into the practices of the field. For instance, Stadler et al. (2019) developed one single procedure to evaluate the effectiveness and efficiency of as well as the user satisfaction with an eHMI concept all at once. Participants were tasked with jaywalking in front of an approaching AV equipped with the interface in a VR environment. In a methodological mix-and-match of sorts, which resulted in confounding the effectiveness of the concept with its efficiency, the design rationale was not explained beforehand, no instruction to "respond as fast and accurately as possible" was given, there was a time limit at the trial level, and error rates, RTs, and NASA-TLX ratings were collected, Similarly, Chang et al. (2018) compared five existing interfaces, developed by automotive manufacturers, technology companies, and research groups, to communicate the intention of an AV to other road users. Participants watched animated videos of an AV equipped with each interface approaching an unsignalized crosswalk and were tasked with making judgments about the AV's intention regarding yielding. In their evaluation procedure, the design rationale was not explained beforehand, instruction to "respond as fast and accurately as possible" was given, there was a time limit at the trial level, and error rates and RTs were collected. Furthermore, Mahadevan et al. (2018) evaluated four interfaces aimed at acknowledging pedestrian presence and signaling AV intention, by measuring participants' crossing intention. In a parking garage, participants were tasked with reporting their intention to cross the street, while a vehicle equipped with one of the interfaces was approaching. In their evaluation procedure, the design rationale was explained beforehand and there was a time limit at the trial level – considering the vehicle travelled a predefined distance at a certain speed while participants contemplated crossing – instruction to "respond as fast and accurately as possible" was not given, and neither RTs nor NASA-TLX ratings were collected.

Having said that, there have been eHMI evaluation attempts that have approximated the evaluation procedure the proposed framework is arguing for. A case in point is Hensch et al. (2019), who evaluated the comprehensibility of an eHMI concept they developed to communicate AV mode and intention to pedestrians. In their study, random pedestrians interacted with a vehicle equipped with the interface in a parking area, and were then asked – among other things – what they thought was indicated by each signal (open-ended question), in an interview that lasted around 5 minutes, providing interviewees with ample time for reflection. In like manner, Ackermann et al. (2019) studied the effect of four interface parameters on eHMI comprehensibility. Participants viewed augmented real-world videos of an AV equipped with an interface approaching and were asked to reflect on the content of what the oncoming vehicle was trying to communicate. It is safe to say that these studies closely approximate the ideal procedure for evaluating eHMI effectiveness. Accordingly, Eisma et al. (2021) studied the effect of an eHMI parameter on crossing decisions, RTs, and eye movements. In their evaluation procedure, instruction to "respond as fast and accurately as possible" was given, there was a time limit at the trial level, and error rates and RTs were collected. Even though the design rationale was not explained beforehand, all the evaluated designs were textual (Walk; Don't walk; Braking; Driving; Go; Stop;) and thus self-explanatory to a great extent. Evidently, this work closely approximates the ideal procedure for evaluating eHMI efficiency.

The proposed framework has inarguably been exemplified in Rouchitsas and Alm (2022; 2023) were the effectiveness and efficiency of an eHMI concept employing facial expressions for communicating AV intention were evaluated in the context of separate evaluation procedures. More specifically, in Rouchitsas and Alm (2022), participants evaluated the effectiveness of said concept without any explanation being provided beforehand regarding the design rationale, no instruction to "respond as fast and accurately as possible", no time limit at the trial level, and with error rates being collected only. On the other hand, in Rouchitsas and Alm (2023), participants evaluated the efficiency of the same concept with clear explanation of the design rationale being provided beforehand, explicit instruction to "respond as fast and accurately as possible", a tight time limit at the trial level, and with RTs complimenting the error rates being collected.

4 DISCUSSION

When evaluating eHMI concepts, it is essential to distinguish between two key usability aspects: effectiveness and efficiency. Effectiveness pertains to whether potential road users will correctly interpret the information conveyed by the eHMI to make appropriate decisions during interactions with AVs. Efficiency, on the other hand, addresses how quickly and effortlessly potential road users can arrive at those decisions in traffic. The question of whether a delayed response is due to a lack of understanding of the intended communication or due to a slower decision-making process highlights the need to disentangle efficiency from effectiveness in eHMI evaluation procedures.

The proposed framework aims to do away with the common and persistent methodological pitfall of confounding the effectiveness of a concept with its efficiency, a pitfall that has plagued the eHMI field since its very inception. The proposed framework manages to accomplish just that by guiding methodological choices regarding design rationale explanation, instructions emphasizing speed, trial-level time limit, and targeted performance measures, depending on whether the research focus is the effectiveness or the efficiency of a given eHMI concept. A clear separation between effectiveness and efficiency ensures a robust evaluation of eHMI concepts, helping researchers and practitioners identify whether issues stem from the communication clarity of the interface or the speed and ease of information processing. Moreover, by employing targeted measures – such as error rates to measure effectiveness and RTs and workload ratings to measure efficiency – researchers and practitioners can better understand the strengths and weaknesses of different eHMI concepts, make valid and reliable comparisons, and proceed with scientifically sound modifications to refine the concepts and ultimately ensure accurate, timely, and effortless responses from road users when interacting with AVs.

5 CONCLUSIONS AND OPEN PROBLEMS

The proposed framework provides a systematic approach to definitively addressing a long-standing methodological issue in the eHMI field, namely disentangling efficiency from effectiveness in eHMI evaluation procedures, and shows great promise for becoming the field's standard evaluation framework for concept development. Nevertheless, the trade-off between effectiveness and efficiency requires further investigation, as the interplay between the two usability aspects can be complicated. Future work should explore cases where improving one might inadvertently compromise the other.

REFERENCES

Ackermann, C., Beggiato, M., Schubert, S., & Krems, J. F. (2019). An experimental study to investigate design and assessment criteria: What is important for communication between pedestrians and automated vehicles? *Applied ergonomics*, *75*, 272–282.

Bazilinskyy, P., Dodou, D., & De Winter, J. (2019). Survey on eHMI concepts: The effect of text, color, and perspective. *Transportation research part F: traffic psychology and behaviour*, *67*, 175-194.

Bevan, N., Carter, J., & Harker, S. (2015). ISO 9241-11 revised: What have we learnt about usability since 1998? In *Human-Computer Interaction: Design and Evaluation: 17th International Conference, HCI International 2015, Los Angeles, CA, USA, August 2-7, 2015, Proceedings, Part I 17* (pp. 143-151). Springer International Publishing.

Calvo-Barajas, N., Rouchitsas, A., & Gürdür Broo, D. (2025). Examining Human-Robot Interactions: Design Guidelines for Trust and Acceptance. In *Human-Technology Interaction – Interdisciplinary Approaches and Perspectives*, Springer.

Chang, C. M., Toda, K., Sakamoto, D., & Igarashi, T. (2017). Eyes on a Car: An Interface Design for Communication between an Autonomous Car and a Pedestrian. In *Proceedings of the 9th international conference on automotive user interfaces and interactive vehicular applications* (pp. 65-73).

Cheema, K., Sweneya, S., Craig, J., Huynh, T., Ostevik, A. V., Reed, A., & Cummine, J. (2023). An investigation of white matter properties as they relate to spelling behaviour in skilled and impaired readers. *Neuropsychological Rehabilitation*, *33*(6), 989-1017.

Dey, D., Habibovic, A., Löcken, A., Wintersberger, P., Pfleging, B., Riener, A., ... & Terken, J. (2020). Taming the eHMI jungle: A classification taxonomy to guide, compare, and assess the design principles of automated vehicles' external human-machine interfaces. *Transportation Research Interdisciplinary Perspectives*, *7*, 100174.

Eisma, Y. B., Reiff, A., Kooijman, L., Dodou, D., & de Winter, J. C. (2021). External human-machine interfaces: Effects of message perspective. *Transportation research part F: traffic psychology and behaviour*, *78*, 30-41.

Hart, S. G. (2006). NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 50, No. 9, pp. 904-908). Sage CA: Los Angeles, CA: Sage publications.

Hensch, A. C., Neumann, I., Beggiato, M., Halama, J., & Krems, J. F. (2019). Effects of a light-based communication approach as an external HMI for

Automated Vehicles--a Wizard-of-Oz Study. *Transactions on Transport Sciences*, *10*(2).

ISO/TR 23049:2018, (2018). *Road Vehicles: Ergonomic Aspects of External Visual Communication from Automated Vehicles to Other Road Users.* London: BSI.

Kantowitz, B. H., Roediger III, H. L., & Elmes, D. G. (2014). *Experimental psychology*. Cengage Learning.

Kaß, C., Schoch, S., Naujoks, F., Hergeth, S., Keinath, A., & Neukum, A. (2020). Standardized test procedure for external Human–Machine Interfaces of automated vehicles. *Information*, *11*(3), 173.

Kyllonen, P. C., & Zu, J. (2016). Use of response time for measuring cognitive ability. *Journal of Intelligence*, *4*(4), 14.

Mahadevan, K., Somanath, S., & Sharlin, E. (2018). Communicating awareness and intent in autonomous vehicle-pedestrian interaction. In *Proceedings of the 2018 CHI conference on human factors in computing systems* (pp. 1-12).

Rouchitsas, A., & Alm, H. (2019). External human–machine interfaces for autonomous vehicle-to-pedestrian communication: A review of empirical work. *Frontiers in psychology*, *10*, 2757.

Rouchitsas, A., & Alm, H. (2022). Ghost on the windshield: Employing a virtual human character to communicate pedestrian acknowledgement and vehicle intention. *Information*, *13*(9), 420.

Rouchitsas, A., & Alm, H. (2023). Smiles and angry faces vs. nods and head shakes: Facial expressions at the service of autonomous vehicles. *Multimodal Technologies and Interaction*, *7*(2), 10.

SAE International J3016. (2021). *Taxonomy and Definitions of Terms Related to Driving Automation Systems for on-road Motor Vehicles.* Available at: www.sae.org (accessed on January 2, 2025).

Schömig, N., Kremer, C., Gary, S., Forster, Y., Naujoks, F., Keinath, A., & Neukum, A. (2024). Test procedure for the evaluation of partially automated driving HMI including driver monitoring systems in driving simulation. *MethodsX*, *12*, 102573.

Stadler, S., Cornet, H., Novaes Theoto, T., & Frenkler, F. (2019). A tool, not a toy: using virtual reality to evaluate the communication between autonomous vehicles and pedestrians. *Augmented Reality and Virtual Reality: The Power of AR and VR for Business*, 203-216.

Wiese, E., Metta, G., & Wykowska, A. (2017). Robots as intentional agents: using neuroscientific methods to make robots appear more social. *Frontiers in psychology*, *8*, 1663.